

AALTO –YLIOPISTON TEKNILLINEN KORKEAKOULU  
Elektroniikan, tietoliikenteen ja automaation tiedekunta

Tero Dubrovin

## **Tilastollinen tekstianalyysi kaupunkisuunnittelun apuvälineenä osallistavassa paikkatietojärjestelmässä**

Diplomityö

Espoo, 1. maaliskuuta, 2010

Valvoja: Professori Olli Simula, TKT

Ohjaaja: Aija Staffans, TKT ja professori Timo Honkela, TKT

AALTO-YLIOPISTO TEKNILLINEN KORKEAKOULU Elektroniikan, tietoliikenteen ja automaation tiedekunta AS-koulutusohjelma		DIPLOMITYÖN TIIVISTELMÄ	
Tekijä: Tero Dubrovin			
Työn nimi: Tilastollinen tekstianalyysi kaupunkisuunnittelun apuvälineenä osallistavassa paikkatietojärjestelmässä			
Sivumäärä: 115		Päiväys: 1.3.2010	
		Julkaisukieli: suomi	
Professori: Informaatiotekniikka		Professuurikoodi: T-115	
Työn valvoja: Prof. Olli Simula, TKT			
Työn ohjaaja(t): Aija Staffans, TKT ja prof. Timo Honkela, TKT			
<p>           Tiivistelmä: Paikallisilla toimijoilla, kuten tietyn alueen asukkailla ja yrittäjillä, on paljon tietämystä omasta toimintaympäristöstään. Tätä paikallistietämystä voidaan hyödyntää kaupunkisuunnittelun päätöksentekoprosessissa käyttämällä tarkoitukseen sopivia käytäntöjä ja teknisiä apuvälineitä, kuten osallistavaa paikkatietojärjestelmää.         </p> <p>           Osallistamisella tarkoitetaan kaupunkisuunnittelussa paikallisten toimijoiden mukaan ottamista suunnitteluun. Osallistamisen korkealla tasolla on myönteisiä vaikutuksia suunnittelun tulosten laatuun, kestävyYTEEN ja julkiseen vastaanottoon sekä kansalaisten voimaantumiseen. Osallistamisen edistämiseksi paikallistietämystä on kuitenkin pystyttävä analysoimaan ja kommunikoimaan päättäjille nykyistä tehokkaammin ja tasapuolisemmin. Tieto on hajautunut lukuisien ihmisten mieliin ja sitä on usein erittäin paljon, joten sen käsitteleminen perinteisillä menetelmillä on vaikeaa ja työlästä.         </p> <p>           Tässä työssä esitetään miten tekstimuotoisena kerättyä paikallistietämystä voidaan analysoida tilastollisen tekstianalyysin menetelmillä ja miten sen puheenaiheita voidaan automaattisesti esittää kartalla alueina siten, että tietoa voidaan hyödyntää tehokkaasti kaupunkisuunnittelussa.         </p> <p>           Kaupunkisuunnittelijoiden ja paikallisten toimijoiden yhteiskäytössä olevan osallistavan paikkatietojärjestelmän osaksi toteutettiin paikallistietämyksen puheenaiheita interaktiivisella kartalla visualisoiva tekstianalyysipalvelu. Työssä käydään läpi suunnittelun teoreettiset ja käytännön lähtökohdat, menetelmän toimintaperiaate, työn tekniset toteutusratkaisut sekä palvelun toiminnan arviointi suhteessa sille asetettuihin vaatimuksiin.         </p>			
Asiasanat: tekstianalyysi, latenti semanttinen indeksointi, osallistava paikkatietojärjestelmä, kaupunkisuunnittelu, visualisointi, Voronoi-diagrammi			

AALTO UNIVERSITY SCHOOL OF SCIENCE AND TECHNOLOGY Faculty of Electronics, Communications and Automation Degree Programme of Automation and Systems Technology		ABSTRACT OF THE MASTER'S THESIS	
Author: Tero Dubrovin			
Title: Statistical language analysis as a tool of urban design in a participatory geographic information system			
Number of pages: 115	Date: 1/3/2010		Language: Finnish
Professorship: Information technology		Code: T-115	
Supervisor: Prof. Olli Simula, D.Sc. (Tech.)			
Instructor(s): Aija Staffans, D.Sc. (Tech.), Prof. Timo Honkela, D.Sc. (Tech.)			
<p>Abstract: Stakeholders have a great amount of knowledge about their living area and environment. This indigenous knowledge may be utilized in a city planning process by using suitable practices and technical tools, such as a participative geographic information system.</p> <p>In city planning, participation means involving the stakeholders in the planning. High level participation enhances the outcome of a planning process by making the results more sustainable and acceptable. It also supports citizen empowerment in general. Indigenous knowledge must be analyzed and communicated more effectively and equally to raise the level of participation. Traditional methods face difficulties in processing the vast amount of knowledge scattered in several minds of local people.</p> <p>This thesis presents a method for analyzing indigenous knowledge by the means of statistical language analysis and geographic visualization. A web service was designed and implemented as a part of a participative GIS that is used by city planners and local operators. The service automatically visualizes topics of knowledge on an interactive map in a way that supports city planning.</p> <p>The thesis includes a review of theoretical and practical bases of the development, the mechanics of the method, technical solutions and the evaluation of the service and its functionality in relation to its requirements.</p>			
Keywords: text analysis, latent semantic indexing, participatory geographic information system, city planning, visualization			

## Alkusanat

Haluan kiittää Yhdyskunta- ja kaupunkisuunnittelun laboratoriota ja koko Arkkitehtuurin laitosta poikkitieteellisestä avoimuudesta sekä siitä, että toivotitte minut tervetulleeksi tekemään diplomityötäni teidän osastollenne.

Haluan kiittää työni ohjaajia Aija Staffansia ja professori Timo Honkelaa erinomaisesta ohjauksesta, huikeasta ammattitaidosta sekä ymmärryksestä monen tiedekunnan välillä seikkailevaa opiskelijaa kohtaan. Valvojaani professori Olli Simulaa haluan kiittää näiden samojen asioiden lisäksi erityisesti opastamisesta käytännön asioiden kanssa. Pilvi Nummille, Tero Alénille ja Kehittämisfoorumin ylläpitäjille myös kiitokset yhteistyöstä ja konsultoinnista.

Kiitokset myös huoltojoukkueelle, ennen kaikkea kärsivälliselle ja luonnonkauniille tyttöystävälleni Ann-Marielle, tien raivanneelle veljelleni Jorille ja siskolleni Tanjalle, aurinkoiselle siskonpojalleni Aapolle, äidilleni Marja-Tertulle sekä ystävilleni, erityisesti Ilkka Törmälle arkkitehtiympyröiden avaamisesta.

Omistan työni isäni Yrjön muistolle.

Espoossa, 1.3.2010

Tero Dubrovin

# Sisällysluettelo

<b>1</b>	<b>JOHDANTO.....</b>	<b>1</b>
1.1	TAUSTAA .....	1
1.2	TYÖN TAVOITTEET JA RAJAUS .....	4
1.3	TUTKIMUSMENETELMÄT .....	4
1.4	TYÖN RAKENNE .....	5
<b>2</b>	<b>TEOREETTISET LÄHTÖKOHDAT.....</b>	<b>6</b>
2.1	OSALLISTAMINEN .....	6
2.1.1	Paikallistietämys .....	9
2.1.2	Verkkodemokratia .....	10
2.1.3	Kritiikkiä verkkodemokratiasta ja osallistamisesta.....	12
2.2	PAIKKATIE TOJÄRJESTELMÄT.....	13
2.2.1	Julkiset interaktiiviset karttapohjaiset verkkopalvelut .....	14
2.2.2	Osallistava GIS .....	16
2.2.3	Kritiikkiä osallistavasta GIS:stä.....	17
2.3	PAIKKATIEDON VISUALISOINTI .....	19
2.3.1	Epämääräisen paikkatiedon visualisointi .....	21
2.3.2	Alueiden raja us .....	22
2.3.3	Puheenaiheiden visualisointi paikkatietojärjestelmässä.....	25
2.4	AUTOMAATTINEN TILASTOLLINEN TEKSTIANALYYSI.....	26
2.4.1	Vektoriavaruusmalli .....	28
2.4.2	Dimension pienennys.....	29
2.4.2.1	Pääkomponenttianalyysi .....	30
2.4.2.2	Riippumattomien komponenttien analyysi .....	31
2.4.2.3	Itseorganisoituva kartta.....	32
2.4.2.4	Muut dimensionpienennysmenetelmät .....	33
2.4.3	Latentti semanttinen indeksointi .....	34
<b>3</b>	<b>KEHITYSKOHDE.....</b>	<b>36</b>
3.1	KEHITTÄMISFOORUMI .....	36
3.1.1	Kehittämisfoorumin käyttö kaupunkisuunnittelussa.....	36
3.1.2	Aineisto .....	37
3.2	JÄRJESTELMÄN TEKNOLOGIAT .....	37
3.2.1	Sisällönhallintajärjestelmä .....	38
3.2.2	Karttapalvelu .....	38
3.2.3	Palvelin .....	39
3.2.4	Tietokanta .....	39
3.3	PALVELUN VAATIMUKSET.....	40
3.3.1	Käyttäjien tarpeet.....	40
3.3.2	Palvelun sopivuus olemassa olevaan järjestelmään .....	42
3.3.3	Vaatimukset osallistavuuden parantamiseksi.....	43
3.3.4	Tekniset vaatimukset .....	45
3.3.5	Julkisen interaktiivisen verkkopalvelun menestystekijät .....	47
<b>4</b>	<b>KEHITETTÄVÄN PALVELUN MÄÄRITTELY .....</b>	<b>49</b>
4.1	PALVELUKOKONAISUUDEN KUVAUS .....	49
4.1.1	Puheenaiheet tietyssä kartan pisteessä .....	50
4.1.2	Samankaltaisten kommenttien tunnistus .....	51
4.1.3	Klusterointi .....	51
4.1.4	Alueen raja us puheenaiheen mukaan .....	52
4.2	TOTEUTETTAVAKSI VALITUN PALVELUN MENETELMÄ.....	53
4.2.1	Tekstianalyysi .....	53

4.2.2	Kommenttien siirto hyperpallon kehälle .....	55
4.2.3	Tekstiaineiston esikäsittely .....	56
4.2.4	Alueen muodostuksessa käytetty menetelmä .....	56
4.3	PALVELUN TOIMINNALLISUUS KÄYTTÄJÄN KANNALTA .....	57
4.3.1	Käyttötapaukset .....	58
4.4	MENETELMÄN ASETTAMAT VAATIMUKSET .....	59
<b>5</b>	<b>TEKNINEN RATKAISU .....</b>	<b>62</b>
5.1	PALVELUN TOIMINNAN LOGIIKKA .....	62
5.2	JÄRJESTELMÄN SISÄINEN KOMMUNIKAATIO .....	64
5.3	KÄYTTÖLIITTYMÄMUUTOKSET .....	65
5.4	TEKSTIANALYYSI .....	66
5.4.1	Sanan katkaisu .....	67
5.5	ALUEIDEN MUODOSTUS .....	67
5.6	VISUALISOINTI .....	68
<b>6</b>	<b>TYÖN ARVIOINTI .....</b>	<b>71</b>
6.1	MENETELMÄN EVALUOINTI .....	71
6.1.1	Tekstianalyysin evaluointi .....	71
6.1.2	Spatiaalisen analyysin evaluointi .....	75
6.2	PALVELUN VAATIMUSTEN TOTEUTUMINEN .....	78
6.2.1	Käyttäjien tarpeiden toteutuminen .....	78
6.2.2	Yhteensopivuus Kehittämisfoorumin kanssa .....	80
6.2.3	Osallistavuuden edistäminen .....	82
6.2.4	Teknisten vaatimusten toteutuminen .....	83
6.2.5	Palvelun menestymisvaatimusten toteutuminen .....	85
6.2.6	Aineiston vaatimukset .....	87
6.3	PALVELUN ARVIOINTI KOKONAISUUTENA .....	88
6.4	TOTEUTETUN PALVELUN JATKOKEHITYS .....	90
6.4.1	Tekstianalyysin vaihtoehtoisia toimintatapoja .....	90
6.4.2	Vaihtoehtoja käytettävän aineiston suhteen .....	91
6.4.3	Alueiden rajaamisen kehitys ryhmittelyn avulla .....	92
6.4.4	Spatiaalinen tekstianalyysi .....	92
6.5	PALVELUKOKONAISUUDEN MUIDEN PALVELUJEN TOTEUTTAMINEN .....	93
<b>7</b>	<b>JOHTOPÄÄTÖKSET .....</b>	<b>95</b>
<b>8</b>	<b>YHTEENVETO .....</b>	<b>97</b>
	<b>LÄHDELUETTELO .....</b>	<b>98</b>

# **Liiteluettelo**

Liite 1. Analyysipalvelun käyttötapaukset

# 1 Johdanto

Tämä diplomityö on osa Opus, Oppiva kaupunkisuunnittelu ja asumisen arki -hanketta, joka on monitieteinen tutkimushanke joka toteutettiin vuosien 2005-2009 välisenä aikana. Hankkeen toteutti kaksi Aalto-yliopiston teknillisen korkeakoulun yksikköä, tietotekniikan osastoon kuuluva liiketoimintaprosessien simulointilaboratorio SimLab ja Arkkitehtuurin laitokseen kuuluvan yhdyskunta- ja kaupunkisuunnittelun laboratorio. Tutkimuksen päärahoittajana oli Tekes. Hankkeessa kehitetään innovatiivisia Internet-pohjaisia työkaluja, joilla täydennetään käytössä olevia osallistavia suunnittelumenetelmiä, sekä korostetaan asukkaiden ja ympäristön muiden toimijoiden asemaa osana alueen kehitysverkostoa. (OPUS Oppiva kaupunkisuunnittelu ja asumisen arki 2010)

Työn osana toteutetaan julkisessa käytössä olevan paikkatietojärjestelmän osaksi palvelu, jonka on tarkoitus edistää kaupunkisuunnittelua kansalaisia osallistavalla tavalla tilastollisen tekstianalyysin ja visualisoinnin menetelmiä hyödyntäen. Palvelu visualisoi automaattisesti kansalaisilta kerättävien kommenttien puheenaiheita interaktiivisella kartalla.

## 1.1 Taustaa

*Kaupunkisuunnittelu* on käytäntö, jolla yhteisön jäsenet pyrkivät edistämään alueensa kehitystä sekä parantamaan elinolojaan. Suunnittelutyö on asiantuntijapainotteista ja siihen käytetään lukuisia erilaisia keinoja ja apuvälineitä. Pääasiassa suunnittelu tapahtuu kaavoituksella, joka on kuntien monopoli ja merkittävä asiantuntijainstituutio, jossa kuitenkin asukkaillakin on osallistumisoikeus. (Staffans 2004)

Kaupunkisuunnittelun tuloksilla on vaikutusta alueella toimivien paikallisten ihmisten, kuten alueen asukkaiden ja yrittäjien, elämään ja toimintaan, joten heillä on intressi olla mukana suunnittelussa. Suomessa Maankäyttö- ja rakennuslaki (Maankäyttö- ja rakennuslaki 1999b) määrää että kaavaluonnos on oltava yleisesti nähtävillä ennen sen hyväksymistä, ja sitä on voitava kommentoida. Laki määrää myös, että kaavoittajan on



suunnittelutyössään kuultava *osallisia* (Maankäyttö- ja rakennuslaki 1999a). Osallisiksi on määritelty alueen maanomistajat sekä ne, joiden asumiseen, työntekoon tai muihin oloihin kaava saattaa huomattavasti vaikuttaa, sekä viranomaiset ja yhteisöt, joiden toimialaa suunnittelussa käsitellään (Maankäyttö- ja rakennuslaki 1999a). Laissa ei kuitenkaan täsmennetä, miten kuulemisen on tapahduttava, tai miten saatua tietoa pitää käyttää. Periaatteessa osallisten mielipiteet voi laillisesti jättää täysin huomioimatta lopullisessa kaavassa, varsinaisia valituksia lukuun ottamatta.

Asukkaiden *osallistamisella* tarkoitetaan kaupunkisuunnittelussa asukkaiden mukaan ottamista kaavojen valmisteluun. Asukkailla on paljon sellaista tietoa omasta asuinympäristöstään, jota ulkopuolisella asiantuntijalla ei luonnostaan ole. Tätä tietoa hyödyntämällä voidaan tehdä paremmin informoituja päätöksiä alueen kehittämisestä.

Ylhäältä käsin suoritettavassa suunnittelussa ei voida jatkuvasti kartoittaa alueen asukkaiden muuttuvia tarpeita, vaan suunnittelu perustuu ennalta valittuihin lähtökohtiin ja malleihin. Perinteistä kaupunkisuunnitteluprosessia on kritisoitu sen kyvyttömyydestä vastata osallisten vaihteleviin tarpeisiin esimerkiksi alueen suojelun ja kehityksen suhteen. Puutteita on erityisesti osallistumismahdollisuuksien *tasavertaisuudessa* (engl. ”equity”) ja *saavutettavuudessa* (engl. ”access”). Saavutettavuudella tarkoitetaan osallisten mahdollisuuksia vaikuttaa suunnitteluun lopputuloksen kannalta merkityksellisellä tavalla, ja tasavertaisuudella sitä, että kaikkien osallisten vaikutusmahdollisuudet ovat yhtä suuret. Saavutettavuus ja tasavertaisuus kärsivät erityisesti suunnittelun aika- ja paikkasidonnaisuudesta. (Dragičević & Balram 2004)

Kansalaisten mahdollisuutta tehdä päätöksiä oman elinympäristönsä suhteen suoraan Internetin välityksellä kutsutaan *verkkodemokratiaksi*. Verkkodemokratialla on ristiriitainen asema kaupunkisuunnittelussa. Aihetta on tutkittu viime aikoina paljon, eikä pidetä itsestään selvänä että se parantaisi kaupunkisuunnittelun tuloksia. Kuitenkin asukkaiden ja kaupunkisuunnittelijoiden välisen kommunikaation parantaminen ja molempipuolisen ymmärryksen lisääminen on verkkodemokratiaan kriittisesti suhtautuvienkin mielestä tarpeellista hyvään lopputulokseen pääsemiseksi. Jo pelkästään se, että ihmiset voivat osallistua päätöksentekoon saadessaan ilmaista mielipiteensä, saa

heidät kokemaan tehtävät päätökset parempina ja oikeudenmukaisempina (Carpini *et al.* 2004) (Ruusula 2001).

*Paikkatietojärjestelmä* (engl. ”Geographic Information System”, lyh. GIS) on interaktiivinen ohjelmistokokonaisuus, jonka avulla käyttäjät voivat käsitellä ja analysoida *paikkatietoa* eli kartalle sijoitettua informaatiota visuaalisessa muodossa. Paikkatietojärjestelmiä on alettu käyttää apuna myös osallistamisessa. *Osallistavan paikkatietojärjestelmän* tarkoituksena on edistää kommunikaatiota ja tehostaa kansalaisten kykyä tehdä aloitteita heitä koskevista asioista. Tämän vaikutuksen aikaansaamiseksi järjestelmän on kuitenkin tarjottava mahdollisuus merkitykselliseen vuorovaikutukseen. Osallistavuuden onnistumista ei voida mitata pelkillä teknisillä saavutuksilla, vaan lopputulokseen vaikuttaa paljon myös sosiaalinen ympäristö ja tavat, joilla uusi teknologia otetaan käyttöön. (Tripathi & Bhattarya 2004) (Dunn 2007)

Aalto-yliopiston teknillisen korkeakoulun Arkkitehtuurin laitoksen Yhdyskunta- ja kaupunkisuunnittelun tutkijaryhmän tuottama ja ylläpitämä *Kehittämisfoorumi*-Internet-sivusto tarjoaa palvelun, jonka avulla tietyn alueen asukkaat voivat kommentoida omaa asuinympäristöään jättämällä paikkatietojärjestelmän kartalle tekstimuotoisia kommentteja haluamiinsa paikkoihin. Kommentteja hyödynnetään kaupunkisuunnittelun tietolähteenä, mutta niiden sisällöstä on kuitenkin työlästä muodostaa kokonaiskuvaa, jonka perusteella kaupunkisuunnittelijat voisivat suunnitella alueen kehittämistä.

*Tilastollisella tekstianalyysillä* voidaan analysoida kommenttien puheenaiheita siten, että ne voidaan esittää paikkatietojärjestelmän kartalla visuaalisessa muodossa. Hyvä visualisointi saa käyttäjän havaitsemaan kokonaisuuden ilman, että hänen tarvitsee yrittää muodostaa kokonaiskuvaa mielessään (MacEachren 1992). Tästä on apua kaavoittajalle, joka yrittää kartoittaa paikallisten toimijoiden tarpeita päätöksenteon tueksi. Visuaalinen esitystapa on kaiken tasoisille käyttäjille helposti ymmärrettävissä (Al-Kodmany 2002), joten analyysipalvelu voidaan tarjota myös kansalaisten itsensä käytettäväksi. Suunnitteluprosessista tulee avoimempi, kun alueen asukkaat ja yrittäjät voivat itse kokeilla, miten heidän antamiaan kommentteja käytetään ja miten ne suhtautuvat kokonaisuuteen.

## **1.2 Työn tavoitteet ja raja**

Tämän diplomityön tavoitteena on selvittää *voidaanko kansalaisten osallistamista ja kaupunkisuunnittelun prosessia parantaa paikkatietojärjestelmässä toteutetun tilastollisen tekstianalyysipalvelun avulla*. Samalla kartoitetaan osallistavuuden merkitystä kaupunkisuunnittelun ja kansalaisten hyvinvoinnin kannalta sekä vaatimuksia sen edistämiseksi erityisesti paikkatietojärjestelmän avulla.

Työn keskeisenä osana on Kehittämisfoorumin paikkatietojärjestelmän osaksi suunniteltava tekstianalyysipalvelujen kokonaisuus, josta työn puitteissa toteutetaan yksi palvelu. Palvelujen tarkoituksena on edistää kansalaisten ja kaupunkisuunnittelijoiden välistä kommunikaatiota siten, että kansalaisten mielipiteistä ja niiden spatiaalisesta luonteesta saa helposti muodostettua totuudenmukaisen kokonaiskuvan.

Vaikka käytettävyyssuunnittelulla voidaan edesauttaa paikkatietojärjestelmän osallistavuutta, työssä ei paneuduta syvällisesti käytettävyyssnäkökohtiin, koska Kehittämisfoorumin käytettävyydestä on aiemmin jo tehty diplomityö (Nummi 2007). Työssä ei myöskään ole tarkoitus kartoittaa kaupunkisuunnitteluprosessin pullonkauloja, vaan painopiste on kansalaisten osallistamisessa.

## **1.3 Tutkimusmenetelmät**

Työn teoriaosuus on tehty pääasiassa aiheeseen liittyvien artikkelien ja kirjallisuuden pohjalta. Se sisältää myös alan asiantuntijoilta seminaarien ja keskustelujen myötä opittua tietoa sekä omaa päättelyä.

Työn kokeellinen osuus on toteutettu ohjelmoimalla Internetissä toimivan paikkatietojärjestelmän osaksi palvelu, joka käsittelee järjestelmän tietokannassa olevaa paikkatietoa käyttäjän antamien kommentojen pohjalta ja näyttää analyysin tuloksen käyttäjän Internet-selaimessa paikkatietojärjestelmän kartalla.

Toteutetun palvelun toimintaa arvioidaan eri artikkeleista kerättyjen yleisten tarpeiden ja vaatimusten sekä koekäyttäjänä toimivan kaupunkisuunnittelijan antaman palautteen

perusteella.

## **1.4 Työn rakenne**

Työn varsinainen sisältö jakautuu teoria- ja käytännön osuuteen. Teoriaosuudessa käsitellään ensin osallistamiseen liittyviä käsitteitä kuten paikallistietämystä ja verkkodemokratiaa. Seuraavaksi tehdään katsaus paikkatietojärjestelmiin ja niiden käyttöön kansalaisia osallistavalla tavalla. Sen jälkeen käsitellään visualisointia erityisesti laadullisen aluemuotoisen paikkatiedon osalta. Seuraavaksi perehdytään automaattisen tekstianalyysin menetelmiin. Teoriaosan lopussa esitetään kuvaus toteutettavan palvelun alustana toimivasta Kehittämisfoorumista ja sen käytöstä kaupunkisuunnittelun osallistavana apuvälineenä, sekä kartoitetaan kirjallisuudesta ja käyttäjiltä kerätyt tarpeet ja vaatimukset osallistavuutta edistävien tekstianalyysipalvelujen toteuttamiseksi.

Työn käytännön osan alussa määritetään palvelukokonaisuus, josta toteutettavaksi valitaan yksi palvelu. Toteutettavan palvelun menetelmä käydään läpi, samoin sen toiminta käyttäjän kannalta. Työn tekninen toteutus kuvataan tämän jälkeen yksityiskohtaisemmin. Työn arviointi tehdään teoriaosassa määriteltyjen vaatimusten valossa evaluointi- ja käyttöttestauksen perusteella. Käytännön osa päätetään jatkokehitysehdotuksiin.

Työn tulosten arviointi kokonaisuudessaan tehdään luvussa Johtopäätökset. Yhteenveto-osiossa kerrataan vielä työn keskeinen sisältö.

## 2 Teoreettiset lähtökohdat

### 2.1 Osallistaminen

*Osallistamisella* tarkoitetaan paikallisten toimijoiden mukaan ottamista suunnitteluprosessiin. Michael McCall jakaa artikkelissaan vuodelta 2004 osallistamisen neljään eri tasoon sen intensiteetin mukaan. (McCall 2004)

Alimmalla tasolla on *tiedon jakaminen*. Vähimmillään tämä voi tarkoittaa sitä, että suunnittelija ilmoittaa osallisille suunnitelman teknisistä yksityiskohdista.

Kommunikaatio voi olla myös kaksisuuntaista, jolloin suunnitteluun käytetään osallisilta saatua tietoa. Keskustelun aiheet on kuitenkin päätetty ennalta, eikä niihin voi juurikaan vaikuttaa.

Seuraava taso on *konsultointi*. Tässä suunnittelijat kartoittavat systemaattisesti osallisten tarpeita tai mielipiteitä tiettyihin ennalta päätettyihin kysymyksiin.

Kolmannella tasolla on *osallisten ottaminen mukaan päätöksentekoon*. He voivat yhdessä asiantuntijoiden kanssa osoittaa uusia ongelmakohtia, joista asiantuntijat voivat aloittaa uusia keskustelunaiheita laajemman mielipiteen selvittämiseksi.

Korkeimmalla eli neljännellä tasolla on *osallisten oma-aloitteinen toiminta*. Osalliset voivat omasta tahdostaan aloittaa uusia keskustelunaiheita, muodostaa niistä päätöksiä sekä toimeenpanna muutoksia. Osallisilla on todellista valtaa asioiden kulkuun, eivätkä yksittäiset asiantuntijat voi muuttaa päätöksiä ilman, että saavat omille ehdotuksilleen kannatusta.

Sherry Arnstein puolestaan jaotteli jo 60-luvulla osallistamisen kahdeksanportaiseen asteikkoon, josta löytyvät myös yllämainitut tasot (Arnstein 1969). Hänen asteikkonsa jatkuu alemmas, jossa tiedottamisen alapuolelta löytyy vielä *terapia-* sekä *manipulaatiotasot*. Terapialla hän tarkoitti toimijoiden uhriaseman tahallista tai tahatonta korostamista siten, että syytä ja ratkaisua ongelmiin haetaan toimijasta itsestään eikä yhteiskunnasta. Manipuloinnilla puolestaan tarkoitetaan menettelyä, jolla vallankäyttäjät aktiivisesti muokkaavat kansalaisten ajatusmaailmaa siten, että

kansalaiset luulevat toiminnallaan olevan vaikutusta yhteisiin asioihin, vaikka todellisuudessa heillä ei olisi mitään valtaa. Näitä molempia tasoja Arnstein kutsuu yhteisnimellä ”*epäosallistaminen*” (engl. ”nonparticipation”).

McCallin toisen ja kolmannen tason, eli konsultoinnin ja osallisten mukaan ottamisen väliin Arnstein asettaisi *tyynnyttelytason*, jossa kansalaisia ikään kuin lepytellään ottamalla yksittäisiä ruohonjuuritason toimijoita mukaan päätöksentekoeleimiin maallikkojäseniksi. Heillä on siten jonkin verran todellista valtaa, mutta vallankäytölle asetetaan yläraja rajoittamalla maallikkojäsenten lukumäärää.

Lisäksi McCallin korkein taso eli oma-aloitteisuus voidaan jakaa Arnsteinin asteikolla kahteen eri portaaseen; *delegoituun valtaan* sekä *kansalaishallintaan* (engl. ”citizen control”). Molemmissa valta on pohjimmiltaan kansalaisilla itsellään, mutta ero on siinä, käytetäänkö valtaa suoraan vai välikäsiä kautta.

McCallin ja Arnsteinin osallistamistasojen vastaavuus on esitetty kuvassa 1.

	McCall	Arnstein
↑ Osallistavuus	Kansalaisten oma-aloitteinen toiminta	Kansalaishallinta
		Delegoitu valta
	Mukaan ottaminen päätöksentekoon	Mukaan ottaminen päätöksentekoon
		Tyynnyttely
	Kansalaisten konsultointi	Kansalaisten konsultointi
	Tiedon jakaminen	Tiedon jakaminen
		Terapia
		Manipulaatio

**Kuva 1. Osallistavuuden tasot McCallin (2004) ja Arnsteinin (1969) mukaan**

Laki antaa kaavoittajalle suuren vallan kaupunkisuunnittelussa. Toisaalta alueen asukkaat voivat vaikuttaa suunnitelman aikatauluun suurestikin. Esimerkiksi Yhdysvalloissa osalliset voivat hidastaa hanketta jopa vuosikymmenillä vaatimalla tarkkoja dokumentteja ympäristö- tai liikennevaikutuksista (Delaney 2000). Pelkkä projektin hidastaminen ei kuitenkaan ole rakentavaa toimintaa alueen kehitystä ajatellen, vaan parempiin tuloksiin päästään osapuolten välisellä yhteistyöllä. Alueen asukkaille kannattaa antaa muitakin vaikutusmahdollisuuksia pelkän projektin hankaloittamisen lisäksi.

Hyvä yhdyskuntasuunnitteluprosessi vaatii suunnittelijoiden ja osallisten välillä

toimivan kommunikaatiokanavan. Kommunikoinnissa ei ole kyse vain kansalaisten informoinnista, vaan tyydyttävään lopputulokseen päästäkseen kaupunkisuunnittelijoiden tulee saada suunnitelmille yleisön hyväksyntä. Osallisia voitaisiin käyttää myös aktiivisesti apuna suunnittelussa sen sijaan, että vain reagoitaisiin heidän esille tuomiinsa ongelmiin. (Tress & Tress 2003)

### **2.1.1 Paikallistietämys**

*Paikallistietämys* (engl. ”local knowledge” tai ”indigenous knowledge”) on tietoa, joka ei ole yleisesti saatavilla, vaan se on hajautunut tietyn ihmisjoukon, kuten alueen asukkaiden tai kulttuurin jäsenten keskuuteen. Tieto on kokemusperäistä tai yhteisön jäseneltä toiselle kulkeutuvaa perimätietoa. Kaupunkisuunnittelussa paikallistietämyksellä tarkoitetaan paikallisten toimijoiden sellaisia tietoja omasta ympäristöstään, joita ulkopuolisilla asiantuntijoilla ei tavallisesti ole (Tripathi & Bhattarya 2004).

Kaupunkisuunnittelussa asiantuntijoina toimivat kaupunkisuunnittelijat. Paikallisia toimijoita taas ovat alueen asukkaat ja yrittäjät. Julkisessa päätöksenteossa suositaan usein asiantuntijoiden tietämystä maallikoiden tietämyksen sijaan. Asiantuntijoilla on etunaan alaan liittyvä koulutus ja työkokemus sekä systemaattinen lähestymistapa. Päätökset vaikuttavat objektiivisemmalta silloin, kun niiden tekoon on käytetty ulkopuolisen asiantuntijan tietoa. Ei ole kuitenkaan syytä olettaa, että ulkopuolisella asiantuntijalla olisi luonnostaan kaikki tieto käytettävissään, vaan suuri määrä oleellista tietoa voi jäädä käyttämättä. Paikallistietämyksen hyödyntäminen on tärkeätä kehitysprojektien menestyksekkään läpiviemisen kannalta (Tripathi & Bhattarya 2004).

Kaupunkisuunnittelijoilla on enemmän kokemusta kaupunkisuunnittelusta yleensä, mutta jokainen suunniteltava alue on kuitenkin erilainen. Eräs tapa selvittää alueen erityispiirteet on kuulla alueen asukkaita ja hyödyntää heidän paikallistietämystään. Etenkin kun tehdään suuria ihmisjoukkoja koskevia päätöksiä, tulee ottaa huomioon heidän vaihtelevat tarpeensa. Vaikka faktatiedot ja tärkeimmät tunnusluvut saataisiinkin kerättyä asiantuntijavoimin, voivat ihmisten henkilökohtaiset tarpeet tällöin jäädä



pimentoon. Suurin tietomäärä saavutetaan, kun yhdistetään asiantuntijoiden tietämys ja paikallistietämys. Siksi on tärkeää tuoda paikallistietämys mukaan päätöksentekoon. Tieto voidaan tuoda joko asiantuntijoiden käyttöön, tai sitä voidaan hyödyntää suoraan päätöksenteon tietolähteenä.

### 2.1.2 Verkkodemokratia

Länsimainen sivistys perustuu pitkälti demokratialle. Paikallistietämyksestään huolimatta kansalaisten vaikutusmahdollisuudet rajoittuvat pitkälti äänestämiseen, mielenosoituksiin ja lobbaamiseen eli suoraan päättäjiin vaikuttamiseen. Kansalaisten tietoa voitaisiin käyttää myös suoraan päätöksenteossa ilman välikäsiä. Jos suora päätöksenteko tehdään Internetin avulla, puhutaan *verkkodemokratiasta* (engl. ”e-democracy” tai ”teledemocracy”). Verkkodemokratia tarjoaa uudenlaisen mahdollisuuden toteuttaa kansanvaltaa. (Ruusula 2001)

Internet on koko olemassaolonsa ajan saattanut ihmisiä yhteen. Internet ei itsessään tarjoa vaikutusmahdollisuuksia, vaan pelkän kommunikaatiokanavan ihmisten välille. Mikään ei toisaalta estä vallankäyttäjiä luomasta Internet-sivustoa, jonka keräämää tietoa käytetään hyväksi päätöksenteossa. Verkkodemokratiasta on kuitenkin kyse vasta, kun kansalaiset tietävät, että heidän toiminnallaan verkossa on todella vaikutusta lopputulokseen. Osallistavuuden korkeimman tason, kansalaisten oma-aloitteisuuden, saavuttaminen on kansan *voimaantumisen* (engl. empowerment) kannalta tärkeää. Voimaantumisella tarkoitetaan tuntemusta omien vaikutusmahdollisuuksien olemassaolosta. Jos päätöksiä tehdään vain ennalta määrätyistä kysymyksistä, jää todellinen valta helposti kysymyksen asettajan käsiin (Arnstein 1969).

Verkkodemokratiaa voidaan vertailla edustukselliseen demokratiaan, jossa kansalaiset valitsevat itselleen kansanvaaleilla edustajat, ja nämä edustajat puolestaan äänestävät päätöksistä (Scrutiny of acts and regulations committee 2003). Verkkodemokratiassa ei ole edustajia, vaan jokainen edustaa itse itseään. Tämä ei suurten ihmisjoukkojen tapauksessa olisi käytännössä mahdollista ilman teknisiä apuvälineitä. Yhteistä edustukselliselle ja verkkodemokratialle on kuitenkin se, että kummassakin mallissa

pyritään estämään se, että kukaan pääsisi sormeilemaan kansalaisten ilmaisemia tarpeita, vaan sen sijaan tietoa pyritään käyttämään mahdollisimman tasapuolisesti.

Erona verkkodemokratian ja asiantuntijoiden toteuttaman paikallistietämykseen perustuvan päätöksenteon välillä on se, että verkkodemokratiassa kansalaiset pääsevät vaikuttamaan asioihin suoraan ilman välikäsiä. Asiantuntija aiheuttaa tiedonkäsittelyyn joko tietämättään tai tiedostaen virhettä omien näkemystensä ja uskomustensa vaikutuksesta. Myös asiantuntijan kapasiteetti on rajallinen, joten suuresta määrästä paikallistietämystä voi olla vaikeata kerätä oleellista osaa päätöksentekoa varten. Automaattinen tietojenkäsittelykin aiheuttaa päätösprosessiin virhettä järjestelmän epätäydellisyyden vuoksi, mutta virhelähde voi tällöin olla tasapuolisempi kaikille osapuolille. (Henrion *et al.* 1991)

Verkkodemokratiassa ja korkean tason osallistamisessa on kyse myös havaitsemisesta. Paikallisten toimijoiden oikeus tehdä itse aloitteita perustuu heidän havaintokykyynsä. Alueen asukkaat ja yrittäjät elävät itse siinä todellisuudessa, johon he voivat vaikuttaa, jolloin he usein myös havaitsevat ongelmakohdat ensimmäisinä. Heillä on yleensä myös jonkinlainen käsitys siitä, miten ongelman voisi ratkaista, vaikkakin heiltä voi puuttua kokemus laajemman mittakaavan toiminnasta. Muutosten toimeenpano nopeutuu huomattavasti, jos alueen asukkaat voivat tehdä aloitteita itse päättämistään aiheista ilman, että heidän tarvitsee odottaa julkista rahoitusta ja päätöksentekoa, tai yrittää muovata mielipiteitään ennalta päätettyihin kategorioihin. Ongelmakategorioiden lyöminen lukkoon etukäteen on sikälikin epätehokasta, että asukkaiden tarpeet voivat muuttua ajan myötä. (Burke *et al.* 2006)

Demokratian takaamiseksi päätösprosessin olisi oltava kaikille osapuolille helposti saavutettava ja tasavertainen. Kaupunkisuunnittelussa ei perinteisesti ole pyritty demokratiaan, mutta sen tavoittelussa on kuitenkin se etu, että toteutus saa yleisöltä paremman vastaanoton, kun heillä kaikilla on ollut oikeus vaikuttaa lopputulokseen (Dragičević & Balram 2004).

### **2.1.3 Kritiikkiä verkkodemokratiasta ja osallistamisesta**

Virtuaaliyhteisöillä on jo tällä hetkellä todellista poliittista valtaa. Esimerkiksi San Diegon asukkaat ovat saaneet Internetin keskustelupalstojen avulla painostettua kaupunginhallituksen rakentamaan suojapaikan kodittomille (Sui 1997). Vaikka verkkodemokratia on joillakin aloilla osoittautunut toimivaksi malliksi (Kangas & Store 2003), kaikki eivät kuitenkaan suhtaudu siihen yhtä positiivisesti.

Steve Carver (2003) kritisoi mallia liiallisesta vastuun antamisesta yleisölle, vaikka ei ole takeita heidän halustaan tai kyvystään kyseisenlaiseen toimintaan. Maallikoilla voi hänen mukaansa olla hyvinkin harhaisia käsityksiä tosiasioista ja siitä, millä tavoin ongelmia voidaan ratkaista. Edes korkeasti koulutettujen ihmisten tiedot eivät kata kaikkea sitä, mitä vieraalla alalla pitäisi ottaa huomioon. Lisäksi osalliset voivat edelleen nähdä omat vaikutusmahdollisuutensa niin pieninä, että he kokevat ettei heidän kannata haaskata siihen voimavarojaan, kuten perehtymiseen kuluva aikaa. Kritiikkiä saa myös ajatus siitä, että päättäjät ja asiantuntijat todella haluaisivat antaa valtaansa maallikoille. Asemaansa tottunut henkilö saattaa haluta mieluummin toimia perinteisellä tavalla eli oman näkemyksensä pohjalta. Carverin mielestä ajatus siitä, että Internet parantaisi demokratiaa, on liian optimistinen.

Kuutta verkkodemokratiaprojektia analysoivassa artikkelissaan vuodelta 1988 Christopher Arterton totesi projektin rahoituksen suuruuden vaikuttavan kansalaisten osallistumiseen. Suuremmalla budjetilla tehdyt projektit keräsivät yhtä poikkeusta lukuun ottamatta enemmän osallistujia. Myös virkamiesten mukana olon havaittiin lisäävän kansalaisten luottamusta projektiin, koska päätöksillä koettiin tällöin olevan enemmän todellista vaikutusta. Artertonin mukaan tällä hetkellä ihmiset kuitenkin osallistuvat mieluummin projekteihin, joiden tavoitteena on vaikuttaa asioihin valitsemiensa edustajien kautta kuin verkkodemokratian keinoilla. Muusta kirjallisuudesta ei kuitenkaan löydy tukea Artertonin tuloksille. (Arterton 1988)

Suomalaisten yleinen luottamus tieteen ja tekniikan kykyyn parantaa demokratiaa on vaakalaudalla. Tekniikan Akateemisten Liiton TEK:n jäsenten, nuorten sekä päättäjien

näkemyksessä nykytieteen kyvystä ratkaista demokratiaan, ihmisoikeuksiin ja tasa-arvoon liittyviä ongelmia asetettiin suomalaisen tekniikan nykytilaa kartoittavassa *TEKBARO2010*-raportissa (Lehtovirta *et al.* 2010) asteikon puolivälin tuntumaan. Kyseinen tutkimus oli tehty sekä vuonna 2007 että vuonna 2009 kullekin vastaajaryhmälle. Ainostaan nuoret vuonna 2007 vastasivat keskimäärin asteikon puoliväliä (”vaikea sanoa”) hieman positiivisemmin. Kaikkien muiden vastaajaryhmäajankohta -parien vastaus oli keskimäärin puoliväliä negatiivisempi, ja teknologian kyky demokratian parantamiseen arvioitiin huonommaksi kuin esimerkiksi kyky ilmastonmuutoksen tai nälänhädän vähentämiseen.

McCall syyttää monia osallistamisprojekteja niiden taipumuksesta käyttää osallistavaa suunnittelua keinona hyväksyttää kansalla päätöksiä, jotka itse asiassa tehdäänkin keskitetysti. Hänen mukaansa osallistavuuden pitäisi sen sijaan kattaa koko suunnitteluprosessi sen toteutusta ja tulosten seuranta myöten. Oleellista on myös miettiä, mikä on osallistamisen tarkoitus kussakin projektissa. (McCall 2004)

Myös osallistavien järjestelmien tavassa välittää tietoa eri osapuolille on havaittu kehittämisen varaa. Perinteiset esitystavat eivät aina kykene vastaamaan laajan yleisön tarpeita. Visualisointimenetelmiä kehittämällä tilannetta voitaisiin kuitenkin parantaa. (Appleton & Lovett 2005)

## **2.2 Paikkatietojärjestelmät**

*Paikkatiedolla* tarkoitetaan kartalle sijoitettua tietoa, joka sisältää koordinaattitiedon lisäksi muuta informaatiota. *Paikkatietojärjestelmä* on ohjelmistokokonaisuus, jonka avulla yksi tai useampi käyttäjä voi käsitellä ja analysoida paikkatietoa visuaalisesti (Rhyne 1997).

Paikkatietojärjestelmä muodostuu tietokannasta, päätelaitteesta sekä joukosta operaatioita, joilla tietoa voi käsitellä. Järjestelmän käyttöliittymänä toimii karttanäkymä ja ohjaustyökalut, joilla käyttäjä voi lähettää järjestelmälle kommentoja näkymän tai paikkatiedon muuttamiseksi. Paikkatietoa säilytetään tietokannassa, josta se tarvittaessa haetaan käyttäjälle näytettäväksi. Tiedonkäsittelyoperaatioilla on mahdollista muokata

tietokannassa olevaa paikkatietoa. (Rhyne 1997)

Paikkatietojärjestelmän avulla paikkatietoa voi luoda, tallentaa, muunnella, visualisoida sekä analysoida (Dunn 2007). Paikkatietojärjestelmää voidaan luonnehtia myös interaktiiviseksi karttatyökaluksi. Sen käyttö muistuttaa perinteisen paperisen kartan käyttöä, mutta se voi tarjota lisäominaisuuksia, kuten karttanäkymän muuttamisen, paikkatiedon käsittelyn ja analysoinnin sekä kommunikoinnin muiden käyttäjien kanssa (Al-Kodmany 2002). Paikkatietojärjestelmät tarjoavat uudenlaisia näkökulmia asioihin, joiden spatiaalinen luonne ei välttämättä muuten helposti hahmottuisi (Tripathi & Bhattarya 2004).

Paikkatietojärjestelmän karttanäkymä perustuu usein *kerroksiin* (engl. ”layer”) (Tripathi & Bhattarya 2004). Eri kerrokset sisältävät erityyppistä informaatiota, ja usein eri tavoilla esitettynä. Esimerkiksi yhdessä kerroksessa voi olla ilmakuvista muodostettu pohjakartta, toisessa graafisesti piirrettyinä pelkät rakennukset ja kolmannessa värialueina esitettynä asukastiheys. Käyttäjä voi yleensä vaihdella eri kerrosten näkyvyyttä, jolloin hän voi helposti vertailla valitsemiaan tietoja, ja toisaalta pitää karttanäkymän riittävän yksinkertaisena ja selkeänä.

### **2.2.1 Julkiset interaktiiviset karttapohjaiset verkkopalvelut**

*Interaktiivisuudella* eli vuorovaikutuksellisuudella tarkoitetaan tietotekniikassa ihmisen ja koneen välistä vuorovaikutusta. Vuorovaikutus ei tapahdu suoraan ihmisen ja teknologian välillä, vaan se tapahtuu aina jonkin rajapinnan eli käyttöliittymän (engl. ”interface”) kautta. Rajapinnan huolellisella suunnittelulla voidaan parantaa teknologian interaktiivisuutta. Interaktiivisuuden kehittämisessä tulee kiinnittää huomiota myös ohjelman käytettävyyteen sekä kognitiivisiin näkökohtiin. (Crampton 2002)

Maantieteellisen visualisointipalvelun interaktiivisuus voidaan jakaa neljään eri tyyppiin vuorovaikutuksen kohteen mukaan. Eri kohteita ovat aineisto, esitysmuoto, aikaulottuvuus ja konteksti. Esimerkkinä käyttäjän vuorovaikutuksesta aineiston kanssa on aineistolle tehtävä analyysi, joka tuottaa tuloksenaan lisää tietoa aineistosta. Esitysmuotoon voi vaikuttaa esimerkiksi vaihtamalla kuvakulmaa, josta aineistoa

tarkastelee. Joitakin aineistoja voidaan tarkastella myös aika-akselilla, esimerkiksi näyttämällä aineiston keräämisprosessi kartalla sekventiaalisesti aikajärjestyksessä. Kontekstuaalista interaktiivisuutta on esimerkiksi linkkien eli siirtymämahdollisuuksien tarjoaminen järjestelmän ulkopuolisiin systeemeihin. (Crampton 2002)

Vuonna 1998 tehdyn tutkimuksen mukaan Euroopan kaupungit tarjoavat asukkailleen lukuisia erilaisia interaktiivisia verkkopalveluita (Mino 2000). Palveluilla on monenlaisia tarkoituksia, kuten uudenlaisten toimintojen ja paikallistietämyksen käytön edistäminen, uusien yrittäjien houkutteleva alueelle, asukkaiden saattaminen yhteen, osallistaminen, kulttuurin ja turismin edistäminen ja koulutus. Internetin on todettu olevan paras alusta interaktiivisille palveluille sen laajasta levinneisyydestä ja helposta saavutettavuudesta johtuen.

Tutkimuksen mukaan paikkatietojärjestelmä on kaupunkien yleisimmin tarjoama interaktiivinen verkkopalvelutyyppi. Jopa 70 %:lla tutkituista 80:stä eurooppalaisesta yli 250 000:n asukkaan kaupungista oli hyöty- tai koekäytössään oma GIS-palvelu. Esimerkiksi virtuaalimaailma ei ollut tässä käytössä läheskään yhtä suosittu.

*Digitaalisella kaupungilla* tarkoitetaan Internetin avulla toteutettua tietoverkkoa, joka on tiiviissä yhteydessä tietyn kaupungin tai kaupunginosan palveluihin ja kehitykseen. Kaupungin suunnitteluun käytettävästä verkkopalvelusta itsestään tulee helposti myös osa kyseistä kaupunkia (Sui 1997). Digitaalisen kaupungin tarkoituksena on tarjota kaupungin palveluita innovatiivisella tavalla, joskin se päättyy usein vain esittelemään pinnallisesti kaupungin parhaita puolia. Toimivan digitaalisen kaupungin kehittämiseen tarvitaan teknisen kehityksen lisäksi monipuolista ammattitaitoa ja asukkaiden osallistamista, ja sen toteutuksen täytyy perustua kaupungin todellisiin ominaisuuksiin, asukkaisiin ja ongelmakohtiin. (Aurigi 2000)

*Mobiili-* eli *kännykkäpaikannuksen* avulla voidaan jokainen kännykkää käyttävä henkilö muuttaa tietolähteeksi, jolla kerätään aineistoa joko yksityiseen tai julkiseen käyttöön. Tietoa voidaan analysoida ja jakaa eri käyttötarkoitusten mukaan. Menetelmää voidaan käyttää ruohonjuuritason havaitsemisen välineenä esimerkiksi terveydenhuollon, kaupunkisuunnittelun, taiteen sekä luonnonvarojen hallinnan aloilla. Esimerkiksi

Brasiliassa tehdyssä tutkimuksessa kännyköitä käytettiin keräämään tietoa ympäristön äänenvoimakkuuden vaikutuksesta elintasoon. Vaikka mobiilipaikannus herättää paljon kysymyksiä tietoturvaan liittyen, moniin näistä ongelmista löytyy tekninen ratkaisu. Menetelmä voidaan yhdistää myös paikkatietojärjestelmään, jolloin puhutaan *mobiilista interaktiivisesta GIS*:stä (engl. ”Mobile Interactive GIS”, lyh. MIGIS). Toistaiseksi reaaliaikaista paikannusta kännykän ja karttasovelluksen avulla on tehty kuitenkin suhteellisen vähän (Ahas & Mark 2005). (Burke *et al.* 2006)

### 2.2.2 Osallistava GIS

*Osallistava GIS* tai *PGIS* (engl. ”Participatory GIS”) on paikkatietojärjestelmään pohjautuva käytäntö, jonka tarkoituksena on palvella suurta yleisöä (Tripathi & Bhattarya 2004). Osallistava GIS kerää käyttäjikseen päättäjiä, suunnittelijoita, paikallisia toimijoita, sekä paikallistietämyksestä yleisesti kiinnostuneita henkilöitä (Tripathi & Bhattarya 2004). Toisin kuin GIS itse, osallistava GIS ei ole pelkkä teknologia, koska sen toiminta on pitkälti kiinni osallistujista ja kontekstista, jossa sitä käytetään (Dunn 2007).

Osallistava GIS antaa kaikille paikallisille toimijoille oikeuden ja mahdollisuuden käyttää yhteistä paikkatietojärjestelmää. Käytännössä tämä tarkoittaa sitä, että henkilöt, joilta tietoa kerätään, voivat myös itse tutkia ja käyttää aineistoa. Sen tärkeimpänä ominaisuutena voidaan pitää sen voimaannuttavaa vaikutusta (Tripathi & Bhattarya 2004). Kansalaisten osallistamisella on havaittu olevan merkittävä vaikutus alueellisten kehittämisprojektien lopputuloksiin (Tripathi & Bhattarya 2004). Internet-pohjaista avointa paikkatietojärjestelmää käyttämällä suunnittelun tuloksista saadaan myös kestävämpiä (Dragićević & Balram 2004).

Osallistava GIS yhdistelee tehokkaasti eri lähteistä peräisin olevaa kvantitatiivista ja kvalitatiivista tietoa (Tripathi & Bhattarya 2004) (Dragićević & Balram 2004). Paikkatietojärjestelmät soveltuvat erinomaisesti paikallistietämyksen keräämiseen ja säilyttämiseen (Tripathi & Bhattarya 2004). Osallistavaa GIS:iä on käytetty sadoissa projekteissa ympäri maailmaa muun muassa yhteisölliseen resurssienhallintaan,

kaupunkisuunnitteluun, jätteenkäsittelylaitosten sijoitteluun, kuljetuksen suunnitteluun, ongelmien priorisointiin, vaaratekijöiden kartoittamiseen, rajakiistojen ennaltaehkäisyyn, naapuristojen välisen yhteisöjen yhteistyön elvyttämiseen ja ympäristönsuojeluun (McCall 2002, 2004) (Dunn 2007) (Tang & Waters 2005) (Higgs 2006).

Paperisen kartan avulla toteutetun osallistavan suunnitteluprosessin etuina ovat yksinkertaisuus, osallistujien kohtaaminen ja halpuus. Osallistajat saavat motivaation kannalta tärkeitä onnistumisen kokemuksia, kun suunnitelmat ovat käsin kosketeltavissa. Yksinkertaisuus voi kuitenkin rajoittaa suunnittelua, eivätkä pelkät kartat välttämättä riitä monimutkaisten projektien tausta-aineiston esittämiseen. Tietokoneella toteutetut suunnitteluvälineet voivat parantaa osallistavuutta, koska niiden avulla yleisö voi tehdä aineistosta tarkempia johtopäätöksiä analyysien ja tehostetun visualisoinnin myötä. PGIS-sovelluksissa pitäisi kuitenkin kiinnittää nykyistä enemmän huomiota vuorovaikutuksellisuuteen, jotta ne saattaisivat ihmiset merkitykselliseen kanssakäymiseen toistensa ja suunnittelukohteen kanssa. (Al-Kodmany 2002)

Osallistava GIS on yritys yhdistää yhteisöllinen kehitys ja paikkatietojärjestelmän teknologia. Se voidaan nähdä myös vastalauseena elitistiselle, pelkästään asiantuntijoiden hallussa olevalle suunnitteluprosessille. Monien paikkatietojärjestelmien käyttäminen vaatii erityisosaamista, mikä rajoittaa niiden käyttäjäkuntaa. Lisäksi niiden käyttöä on usein rajoitettu tarkoituksellisesti vain tietyn ammattiryhmän keskuuteen. Osallistava GIS puolestaan halutaan tarjota mahdollisimman laajalle yleisölle. Sen vuoksi sen käytön on oltava myös helpompaa kuin asiantuntijajärjestelmien. (Abbot *et al.* 2008)

### **2.2.3 Kritiikkiä osallistavasta GIS:stä**

Usein oletetaan, että osallistava paikkatietojärjestelmä pystyy samaan aikaan esittämään suunnitteluun liittyvät tiedot, vastaamaan maantieteellisiin kysymyksiin sekä osoittamaan ja tyydyttämään kaikkien osapuolten piilevät tarpeet (McCall 2002). Osallistava GIS on kuitenkin saanut osakseen myös paljon kritiikkiä.



Vuonna 1998 Durhamin yliopistossa pidetyssä tilaisuudessa käytiin läpi lukuisien PGIS-projektien toteutumista ja tuloksia (Jordan 1999). Projektien osallistavuuden todettiin olevan vielä lapsenkengissään, ja sen määrittelyssä, ymmärtämisessä ja toteutuksessa nähtiin paljon parantamisen varaa. Vaikka projekteissa mukana olleet asiantuntijat pitivät osallistavuuskysymyksiä tärkeämpinä kuin teknisiä seikkoja, osa projekteista ei kerännyt lainkaan osallistujia. Joissakin projekteista paikallistietämystä käytettiin hyvin pinnallisella tavalla, tai sen sijaan päädyttiin käyttämään vain muita tietolähteitä.

Osallistamisen tavoite ei toteudu, jos osallistavaa päätösprosessia ei ole suunniteltu alueen tarpeet huomioon ottavalla tavalla. Tällöin paikalliset toimijat voivat kokea etääntyvänsä päätösprosessista. Kaikkien ihmisryhmien huomioiminen on usein vaikeata, ja osallistamisprojekti voi aiheuttaa voimaantumisen sijaan jopa syrjäytymisen tunnetta. (Tripathi & Bhattarya 2004)

Koska tiedolla on monia lähteitä, aineiston laatua on vaikeampi valvoa kuin tavallisen paikkatietojärjestelmän aineistoa. Sen vuoksi osallistavan GIS:n paikkatieto on usein epätarkempaa. Tietoa voidaan myös tulkita väärin, sen käyttö voi olla pinnallista tai se voi jäädä kokonaan hyödyntämättä. Lisäksi joitakin tietotyyppejä, kuten maanomistustietoja, voidaan jopa väärinkäyttää. (Tripathi & Bhattarya 2004) (McCall 2004).

Paikkatietojärjestelmän käyttäminen kansalaisten osallistamiseen on suhteellisen uusi, 90-luvulla aloitettu käytäntö, joka voidaan nähdä vastalauseena paikkatietojärjestelmien perinteiselle asiantuntijavaltaa korostavalle käytölle. Paikkatietojärjestelmän osallistavuuden tutkimus liittyy vääjäämättä myös feministiseen maantieteeseen, jonka perustavia kysymyksiä ovat "Kuka hyötyy maantieteellisestä tutkimuksesta?" ja "Kuka omistaa maantieteellisen tutkimuksen?". Feministinen maantiede on osoittautunut tehokkaaksi muun muassa naisten työolojen, oikeuksien ja tarpeiden kuvaamisessa laajassa mittakaavassa. Osallistavien paikkatietojärjestelmien tutkimukseen feministinen maantiede tuo mukanaan arvokkaita näkemyksiä, kuten että laadullista ja määrällistä informaatiota voidaan hyvin tutkia samassa järjestelmässä, ja että paikkatietojärjestelmän avulla voidaan reflektiivisesti tutkia myös järjestelmän itsensä

ominaisuuksia sekä sen kontekstia, mihin tosin usein päädytään tahtomatta tutkimuksen varsinaisen tarkoituksen eli paikallistietämyksen kartoittamisen kustannuksella. (Dunn 2007) (Pain 2004)

Vuonna 2000 tehdyssä tutkimuksessa kokeiltiin osallistavan GIS:n käyttöä maankäytön suunnitteluun San Blas:ssa Meksikon rannikolla (Bojórquez-Tapia *et al.* 2001). Tutkimuksessa pyrittiin kartoittamaan maa-alueiden ominaisuuksia paikallisten maanviljelijöiden, vesiviljelijöiden, karjankasvattajien ja ympäristönsuojelijoiden avulla. Tutkimuksessa havaittiin, että paikallisten toimijoiden oli vaikeaa käyttää osallistavaa GIS:iä johtuen heidän puutteellisesta koulutustaustastaan. Osalla käyttäjistä oli vaikeuksia hahmottaa asioiden maantieteellistä laajuutta. He eivät nähneet eroa paikallisten ja alueellisten ongelmien välillä, vaan katsoivat asioita vain omasta näkökulmastaan. Lisäksi havaittiin, että joitakin asioita, kuten korruptiota tai taloudellisia ongelmia, ei voitu mielekkäällä tavalla kuvata maantieteellisesti.

Tutkimuksessa korostui asiantuntijoiden rooli. Käyttäjät pystyivät tulkitsemaan aineistoa vain asiantuntijoiden teknisellä opastuksella, ja aineiston analysointiin kykenivät vain asiantuntijat itse. Koska asiantuntijat voivat helposti tietoisesti tai huomaamattaan vaikuttaa tulosten tulkintaan, ei voida puhua kansalaisten oma-aloitteisesta päätöksenteosta tai kansalaishallinnasta, vaan tämän tyyppinen osallistaminen jää McCallin osallistamisasteikolla toiselle tai kolmannelle portaalle, eli konsultoinnin tai osallisten mukaan ottamisen tasolle. Osallistavan paikkatietojärjestelmän käyttö ei siis välttämättä johda korkeaan osallistamisen tasoon.

Puutteistaan huolimatta paikkatietojärjestelmä on tehokas alusta välittämään paikallistietämystä (Elwood 2006). Osallistava GIS kytkeytyy positiivisesti hyvään hallintotapaan ja kansalaisten väliseen mielekkääseen dialogiin. (McCall & Minang 2005)

## **2.3 Paikkatiedon visualisointi**

*Visualisoinnilla* tarkoitetaan mielikuvien muodostamista asioiden näkymättömistä suhteista käyttämällä tähän tarkoitukseen sopivia kuvallisia välineitä. Hyvä visualisointi

saa katsojan havaitsemaan esitetyt hahmot ilman, että hänen tarvitsee erikseen yrittää muodostaa niistä kuvaa mielessään. (MacEachren 1992)

Visualisoinnilla on pitkä historia, joka ulottuu sivistyksen alkulähteille saakka. Tietokoneella toteutettu visualisointi pohjautuu pitkälti niihin esitystapoihin, joita on käytetty jo ennen tietokoneiden yleistymistä. Useimmat tutut elementit, kuten värit, viivat, kuvat ja kerrokset soveltuvat hyvin nykyaikaistenkin teknologioiden perustaksi. Visualisointimenetelmien kehittämällä voidaan laajentaa kommunikoinnin mahdollisuuksia ja lisätä sen tehokkuutta. Laadukkaaseen visualisointiin tarvittavat tietokoneet ja ohjelmat löytyvät jo useimpien ihmisten kotoa (Danahy 1999). (Al-Kodmany 2002)

Taulukko- tai tekstimuotoinen tieto kannattaa muuntaa graafiseen muotoon ennen kuin sitä esitetään asiaan perehtymättömille henkilöille (MacEachren 1992). Visuaaliselle kommunikoinnille on ominaista sen havainnollisuus kaikille kohderyhmille. Sen vuoksi sitä on suosittava, kun halutaan välittää samaa informaatiota sekä ammattilaisille että maallikoille. (Al-Kodmany 2002)

Kaupunkisuunnittelussa visualisoinnin on havaittu edistävän kommunikointia kaavoittajien ja yleisön välillä (Tress & Tress 2003). Suunnitteluprosesseja ja eri vaihtoehtojen seurauksia visualisoidaan asianosaisille sekä staattisilla että interaktiivisilla malleilla (Delaney 2000).

Paikkatiedon visualisointi ei ole yleisesti kovinkaan pitkälle kehittynyttä, koska tietokonegrafiikka ja informaation käsittely ovat kehittyneet erillään toisistaan. Viime aikoina eri teknologioita on kuitenkin alettu hyvällä menestyksellä yhdistellä. (Rhyne 1997)

Paikkatietojärjestelmää suunniteltaessa on tärkeää ottaa huomioon, mihin tarkoitukseen ja mille yleisölle se on tarkoitettu. Eri paikkatietojärjestelmillä on myös erilaisia visualisointitarpeita. Suunnittelijoilla täytyy olla hyvä ymmärrys siitä, mitkä työkalut sopivat mihinkin tarkoitukseen ja ympäristöön (Al-Kodmany 2002). Esimerkiksi ammattimaisessa tutkimuskäytössä olevan järjestelmän on tarjottava enemmän erilaisia

toimintoja kuin järjestelmän, jonka tehtävänä on vain välittää informaatiota suurelle yleisölle (MacEachren 1992).

### **2.3.1 Epämääräisen paikkatiedon visualisointi**

Paikkatiedolla voidaan kuvata konkreettisia asioita, kuten rakennelmien sijainteja tai maaston ominaisuuksia, tai abstraktia tietoa, kuten suunnitelmia tai erilaisten tutkimusten tuloksia. Erityisesti abstraktiin tietoon liittyy usein epävarmuutta.

Epävarmuus voi johtua esimerkiksi tiedon tilastollisesta luonteesta, tiedon hankintamenetelmän epävarmuudesta, tai tietolähteen epäluotettavuudesta. Ihmisillä on taipumusta suhtautua luottavaisesti paikkatiedon oikeellisuuteen, vaikka siinä olisikin suurta epävarmuutta (MacEachren 1992).

Yksittäisen paikkatiedon sisältämä informaatiomäärä kasvaa, kun siihen lisätään tieto tiedon epävarmuudesta. Tämän lisätiedon esittämiseksi tarvitaan jokin graafinen ominaisuus, jota vaihtelemalla paikkatiedon epävarmuuden määrä voidaan välittää katsojalle. Lisäksi epävarmuus voi olla erityyppistä riippuen tiedon esittämiseen käytettävästä asteikosta (MacEachren 1992). Maantieteellisen visualisoinnin uutuudesta johtuen ei ole olemassa vakiintunutta käytäntöä siitä, miten tiedon epävarmuus tulisi esittää.

Jackues Bertin määrittä vuonna 1967 kirjassaan *Semiologie Graphique* (Bertin 1967) seitsemän graafista perusmuuttujaa; paikka, koko, kirkkaus (engl. "value"), tekstuuri, väri (engl. "hue"), orientaatio ja muoto. Näistä erityisesti koko ja kirkkaus sopivat hyvin numeerisen informaation, kuten intervalli- tai suhdelukuasteikollisen muuttujan epävarmuuden kuvaamiseen, kun taas väri, muoto ja mahdollisesti orientaatio soveltuvat nominaalisen, ja siten myös ordinaalisen informaation epävarmuuden esittämiseen. Eri tekstuureja puolestaan on suositeltu käytettävän vain ilmaisemaan yksinkertaisesti onko tieto riittävän tarkkaa vai ei, ilman välimuotoja. (MacEachren 1992)

Bertin kokosi listan intuitiivisella tavalla, eikä se perustunut kognitiivisiin tai havaintopsykologisiin tutkimuksiin. Sen vuoksi listaa ei tarvitse pitää kattavana, vaan voidaan tutkia myös muita vaihtoehtoja informaation esittämiseen graafisessa muodossa.

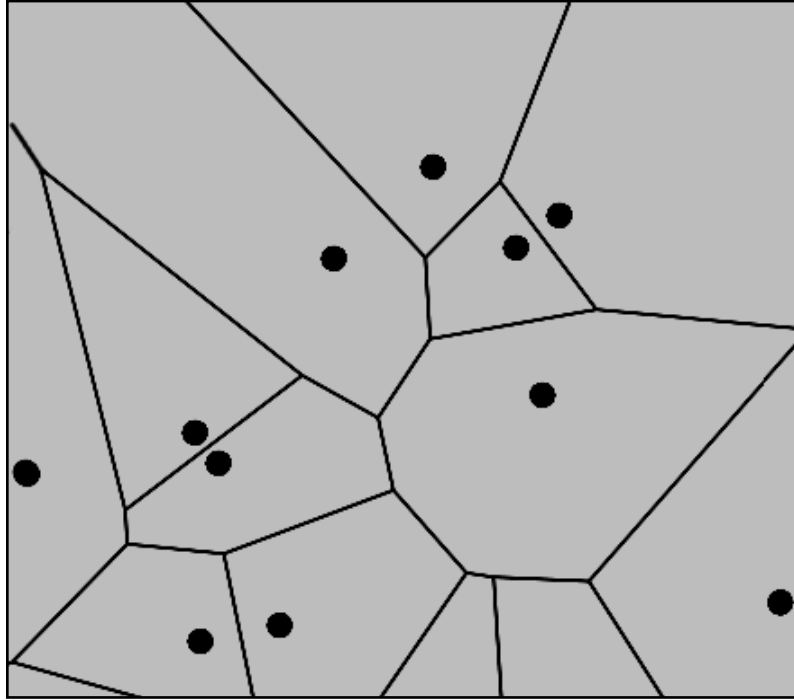
Bertinin listaan kuulumattomista graafisista muuttujista lupaaviksi epävarmuuden esittäjiksi on ehdotettu värikylläisyyttä (engl. ”saturation”), tarkennusta (engl. ”focus”) ja läpikuultavuutta. Värikylläisyydellä tarkoitetaan värin puhtautta, eli sitä, kuinka kaukana väri on harmaasta väristä, ja tarkennuksella hahmon teräväpiirteisyyttä. (MacEachren 2001)

Yksittäisen muuttujan käytön sijaan voidaan käyttää myös yhdistelmää useammasta eri muuttujasta. Yhdistelmän huolellisella valinnalla voidaan vahvistaa mielikuvaa tiedon epävarmuudesta. (MacEachren 2001)

### **2.3.2 Alueiden raja**

Spatiaalinen epätarkkuus on paikkatiedolle tyypillinen epätarkkuuden tyyppi. Spatiaalista epätarkkuutta voi syntyä esimerkiksi, kun maaston ominaisuuksia kuvataan yksittäisistä pisteistä mitatuilla arvoilla. Tällöin ei voida varmuudella tietää, mitä arvoja pisteiden välisessä maastossa on. Usein ei ole mielekäästä ajatella aineiston arvojakaumaa pisteiden kohdalle osuvina piikkeinä, vaan todellisuutta kuvaa paremmin esitysmuoto, jossa paikkatieto visualisoidaan kartalla pisteiden ympärille muodostettuina alueina. Tällöin korostuu tiedon spatiaalinen epätarkkuus ja käyttäjä havaitsee hahmon helpommin yhtenäisenä jatkumona yksittäisten pisteiden sijaan. Kunkin pisteen tiedon ajatellaan siis kuvaavan jossakin määrin myös pisteen ympäristöä. (MacEachren 1992)

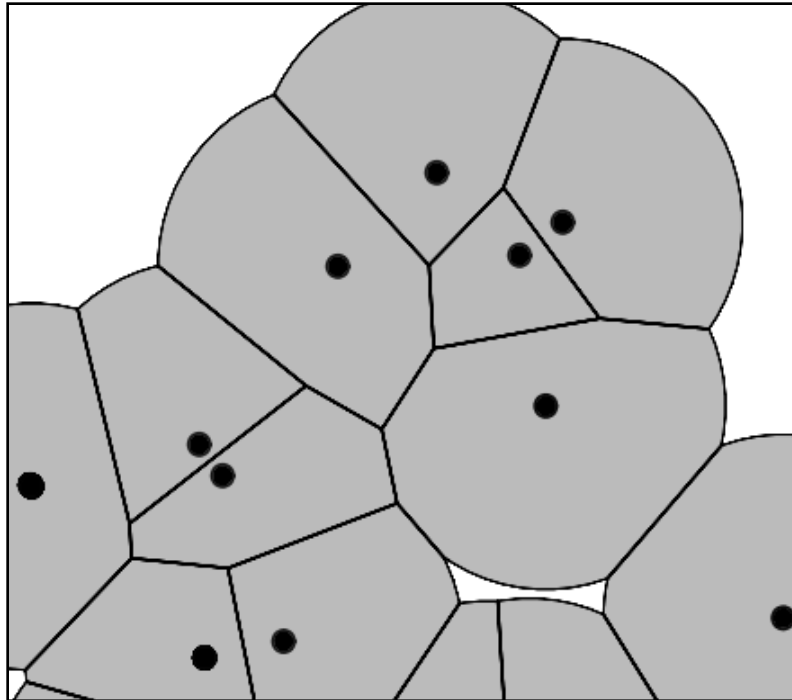
Maaston jakaminen siinä olevia yksittäisiä pisteitä vastaaviksi alueiksi voidaan tehdä *Voronoi-diagrammilla* (engl. ”Voronoi diagram”), joka on yksi hyödyllisimmistä laskennallisen geometrian tietorakenteista (Chew 1990). Voronoi-diagrammin kehitti ukrainalaissyntyinen Georgi Voronoi vuonna 1908. Kaksiulotteisessa tapauksessa Voronoi-diagrammi jakaa maaston alueisiin siten, että kukin maaston kohta kuuluu sitä lähimmän pisteen alueeseen. Tällöin kahden alueen välinen raja kulkee aina vierekkäisten pisteiden puolesta välistä kohtisuorassa pisteitä yhdistävää janaa kohtaan. Kuvassa 2 on esimerkki Voronoi-alueiden muodostumisesta satunnaisten pisteiden tapauksessa. (Fortune 1992)



**Kuva 2. Voronoi-alueiden muodostuminen pisteiden ympärille**

Voronoi-alueet jatkuvat äärettömän kauas peittäen koko maaston. Kun halutaan visualisoida yksittäisten paikkatietopisteiden vaikutusalueita, on järkevää asettaa yläraja sille, kuinka kauas alue ulottuu pisteestä. Sovelluskohteesta kuitenkin riippuu, kuinka suuri etäisyys on sopiva.

Etäisyysrajoitusta voidaan soveltaa Voronoi-alueisiin käyttämällä pisteiden vaikutusalueina Voronoi-alueen ja pisteenkeskisen ympyrän leikkausta. Tällöin pisteen alueeseen valitaan kuuluviksi maastonkohdat, jotka ovat lähempänä kyseistä pistettä kuin yhtäkään muuta pistettä, ja jotka ovat lisäksi riittävän lähellä pistettä. Kuvassa 3 on esitetty pisteiden Voronoi-alueet ympyröillä leikattuina. Kuvan valkoiset alueet eivät kuulu yhteenkään alueeseen.



**Kuva 3. Voronoi-alueet ympyröillä leikattuina**

Voronoi-alue ja ympyrä ovat *konvekseja* eli *kuperia kuvioita*. Konveksisuus voidaan määritellä siten, että jos pisteet A ja B ovat konveksin kuvion sisällä, kaikki pisteet janalla AB ovat myös kuvion sisällä. Määritelmästä seuraa, että Voronoi-alueen ja ympyrän leikkaus on myös konvekksi, koska jokainen piste janalla AB, missä A ja B ovat sekä Voronoi-alueen että ympyrän sisällä olevia pisteitä, on myös molempien kuvioiden, ja siten myös niiden leikkauksen, sisällä. (Anon. 2010)

Rob Ingmanin ja Steve Benfordin vuonna 1995 kehittämässä *LEADS* (LEgibility for Abstract Data Spaces) –systeemissä, jolla pyritään helpottamaan moniulotteisen informaation luettavuutta alueiden avulla, päädyttiin samankaltaiseen ratkaisuun. Aineistosta löytyvien klustereiden välinen raja asetettiin kulkemaan rajan tuntumassa olevien eri luokkiin kuuluvien näytteiden puolesta välistä aivan kuten Voronoi-diagrammissakin. (Ingram & Benford 1995)

### 2.3.3 Puheenaiheiden visualisointi paikkatietojärjestelmässä

Paikallisten toimijoiden toimintaympäristöön kuuluvista paikoista kertovien kommenttien puheenaiheet ovat abstraktia paikallistietämystä. Kommentit sisältävät sijainnin lisäksi myös muuta informaatiota, joten ne ovat myös paikkatietoa.

Kommentin puheenaihe on epämääräistä tietoa, koska se voidaan tulkita monella tavalla. Esimerkiksi lauseen ”*Tämä risteys on vaarallinen.*” voidaan ajatella kertovan muun muassa liikenteestä, liikenneturvallisuudesta, risteyksistä tai niistä kaikista. Usein on myös epäselvää, liittyykö kommentin aihe vain yksittäiseen maastonkohtaan vai laajempaan alueeseen. Esimerkiksi lause ”*Tällaisia grillikatoksia pitäisi olla enemmän.*” kertoo toisaalta vain kyseisestä grillikatoksesta, mutta toisaalta myös laajemmin alueesta, johon grillikatoksia halutaan lisää. Sen vuoksi kommentin puheenaiheen epämääräisyydellä on myös spatiaalinen luonne.

Jos kartalla olevien pistemäisten kommenttien puheenaiheita kuvataan korostamalla tiettyyn puheenaiheeseen liittyvien kommenttien sijainnit, katsoja joutuisi itse mielessään muodostamaan niistä alueen, jolla kyseinen puheenaihe on relevantti. Kappaleessa 2.3 taas todettiin, että hyvä visualisointi saa käyttäjän havaitsemaan esitetyn hahmon ilman ponnistelua. Siksi käyttäjälle on syytä esittää puheenaihe aluemuodossa.

Eräs tapa visualisoida kommenttien puheenaiheita kartalla on käyttää kappaleessa 2.3.2 kuvattua menettelyä, jolla spatiaalinen epätarkkuus tulkitaan jatkamalla kommentin aihealuetta sitä ympäröivään maastoon. Ympäröivät kommentit rajaavat Voronoi-aluetta siten, että maastonkohta kuuluu aina sitä lähimmän kommentin alueeseen.

Toistaiseksi kirjallisuudesta ei löydy tietoa laskennallisista menetelmistä, joiden tarkoituksena olisi visualisoida puheenaiheita alueina paikkatietojärjestelmän kartalla.

Fraunhofer institute AIS:ssä toteutettu paikkatietojärjestelmän ja keskustelupalstan yhdistävä järjestelmä mahdollisti kommenttiaineiston kevyen analysoinnin.

Kommenttien puheenaiheet oli kuitenkin jaoteltu valmiisiin kategorioihin, eikä niiden



epämääräisyyttä pyritty tuomaan esiin visualisoinnin menetelmillä. (Voss *et al.* 2004)

Marketta Kytän ja Maarit Kahilan *PehmoGIS*-tutkimuksessa korostetaan asukkailta kerättävän paikallistietämyksen spatiaalista luonnetta. Tätä kokemuksellista paikkatietoa kuvaillaan ”*pehmeäksi*” paikkatiedoksi. Tietoa on kerätty Internetissä toimivalla palvelulla kuuden kunnan alueella sekä laadullisessa että numeerisessa muodossa. Tutkimuksessa todettiin laadullisen aineiston keräämisen olevan houkuttelevaa, mutta sen analysoimisen työlästä. Aineistoa on pyritty analysoimaan mm. jaotteleamalla kommentteja kiinteisiin aihekategorioihin. (Kytä & Kahila 2007)

Kytän & Kahilan *PehmoGIS*-tutkimus yllyttää tutkimaan keinoja, joilla spatiaalista paikallistietämystä voidaan välittää tehokkaasti päätöksenteon tueksi:

*”PehmoGIS-menetelmillä tuotetun palautetiedon kytkeminen osaksi suunnittelua ja päätöksentekoa edellyttää kuitenkin lisätutkimusta siitä, mitkä ovat joustavimpia tapoja siirtää, päivittää ja ylläpitää tietoa suunnittelijoille, sekä sitoa uutta tietoa kiinteäksi osaksi tämänhetkisiä asiantuntijajärjestelmiä. Tarkoituksena on saada käyttöön suunnittelua tukevia menetelmiä tai järjestelmiä, jotka houkuttelevat asukkaita osallistumaan, toimivat kanavina eri toimijoiden välillä ja tuottavat suunnittelussa ja päätöksenteossa tarvittavaa relevanttia tietoa.”* (Kytä & Kahila 2007)

## **2.4 Automaattinen tilastollinen tekstianalyysi**

Paikallistietämyksen analysoiminen on vaativaa. Tieto on hajautunut eri ihmisten mieliin ja jokaisella se on jäsentynyt eri tavalla. Jos kaikki tutkimuskohteeseen liittyvä tieto kerätään yhteen paikkaan esimerkiksi pyytämällä ihmisiä kirjoittamaan aiheesta ja

keräämällä kirjoitukset, tekstiä tulee helposti valtavan paljon. Sen läpikäyminen ihmisvoimin on pitkä ja uuvuttava prosessi, koska samalla on yritettävä muodostaa ja pitää mielessään kokonaiskuvaa jo luetusta tiedosta. Tekstit pitäisi lukea moneen kertaan, jotta kokonaiskuvasta ei tulisi kovin vääristynyt. Tekstien lukemisjärjestyksellä on vaikutusta siitä jäävän muistikuvan vahvuuteen, mikä vaikuttaa tekstien saaman suhteellisen huomion määrään (Murdock 1962).

Automaattisella tilastollisella tekstianalyysillä voidaan analysoida tekstiaineisto nopeasti ja tasapuolisesti. Tekstin sisältöä ei voida koneellisesti täysin ymmärtää, koska tekstiin sisältyy kielipin lisäksi loputtomasti kulttuurisia elementtejä, joita kaikkia ei voida opettaa tietokoneohjelmalle (Liu *et al.* 2005). Jo pelkkä tekstin kielellinen rakenne on usein monitulkintainen, joten tekstin sisällöstä voidaan tehdä vain tilastollisia tulkintoja. Jos jonkin ohjelman siis väitetään selvittävän lauseen sisällön, se tekee sen oikein vain jollakin todennäköisyydellä. Asiayhteydestä riippuu, mikä lauseen todellinen merkitys on ollut. Edes ihminen ei pysty täydellä varmuudella tulkitsemaan tekstiä oikein, vaikka ihmisellä onkin erinomainen kyky hyödyntää kontekstista saatuja vihjeitä tekstin ymmärtämisessä (Popping 2000).

Koska tekstin täydellistä ymmärtämistä ei voida saavuttaa, tavoitteeksi on otettava jotakin helpompaa. Mahdollisia analyysin tavoitteita ovat esimerkiksi tekstin aihepiiriin selvittäminen, kirjoitustyylin analysointi, kirjoittajan tunnistaminen tai kielen oikeellisuuden tarkistus. Analyysin tavoitteesta on kiinni, mitä sen kohteeksi kannattaa valita. Jos ollaan esimerkiksi selvittämässä tekstin yleispiirteitä kuten puheenaihetta, pelkkä sanastoon kohdistuva analyysi on riittävä. Kielipin tarkistamiseen puolestaan tarvittaisiin sanaston lisäksi myös tietoa kielen rakenteesta.

Osma Suominen tutki vuonna 2008 Pro gradu –tutkielmassaan moninäkömahaun ja tiedon luokittelun menetelmiä. Aineistona käytettiin tekstiartikkeleita ja tietoa siitä, mihin aiheisiin kukin artikkeli liittyy. Tutkielman osana kehitetyllä menetelmällä voitiin selailla aineistoa rajaamalla näkymä eri aihesanojen mukaan. Käytettävyytsteissä kuitenkin ilmeni, että käyttäjät eivät käyttäneet rajausominaisuutta koska se jakoi materiaalin epärelevantilla tavalla, ja koska haun ulkopuolelle jäi artikkeleita, jotka

sinänsä liittyivät valittuun aiheeseen, mutta joihin ei syystä tai toisesta oltu kirjattu kyseistä aihe sanaa. (Suominen 2008)

Automaattinen tilastollinen tekstianalyysi voidaan toteuttaa siten, että sanakategorioita ei päätetä etukäteen, vaan kaikki aineiston sanat toimivat ikään kuin omina kategorioinaan. Sanat voidaan järjestää automaattisesti sen mukaan, kuinka paljon ne liittyvät toisiinsa. Tällöin aineistosta voidaan löytää mihin tahansa sanaan enemmän tai vähemmän liittyvät artikkelit *ohjaamattomasti* eli ilman opetusaineistoa, jossa aineisto olisi etukäteen luokiteltu (Greene *et al.* 2008).

### 2.4.1 Vektoriavaruusmalli

Tilastollisen tekstianalyysin tutkimuksessa on havaittu, että monet laskennalliset menetelmät soveltuvat hyvin tekstiaineiston käsittelyyn, mikäli tekstiaineisto muunnetaan ensin numeeriseen muotoon. Eräs yleinen tekstianalyysin toteutustapa on käsitellä tekstiaineistoa pistejoukkona moniulotteisessa avaruudessa, jonka jokainen ulottuvuus vastaa yhtä aineistosta löytyvää sanaa. Tätä mallia kutsutaan nimellä *vektoriavaruusmalli* (engl. ”vector space model”) tai sanapussimalli (engl. ”bag of words”). ja kyseisestä avaruudesta käytetään tässä työssä nimitystä *sana-avaruus*. Sana-avaruus voidaan esittää matriisina, jonka sarakkeet vastaavat avaruuden ulottuvuuksia ja rivit näytteitä. Jokainen analysoitava tekstinäyte sijoitetaan sana-avaruuteen sen sisältämän sanaston perusteella esimerkiksi siten, että jos näytteestä löytyy tietty sana, näyte saa sanaa vastaavassa ulottuvuudessa arvon 1 ja muussa tapauksessa arvon 0. Voidaan havaita, että tällöin näytteet, joiden sanasto on keskenään samankaltainen, sijaitsevat sana-avaruudessa euklidisessa mielessä lähellä toisiaan. (Bingham 2003)

Sana-avaruudelle tehdään yleensä dimension pienennys, eli vähennetään sen ulottuvuuksien lukumäärää. Tähän on neljä syytä. Ensinnäkin useimmat laskentaoperaatiot ovat helpompia ja nopeampia pienempiulotteisessa avaruudessa. Toiseksi tiedon tallentaminen ja siirtäminen on helpompaa, koska pienemmässä avaruudessa sitä on yleensä vähemmän. Kolmanneksi, jos dimensiota pienennetään riittävästi, aineisto voidaan visualisoida esimerkiksi kaksi- tai kolmiulotteisessa

esitysmuodossa.

Neljäs ja tekstianalyysin kannalta tärkein syy on se, että informaation oletetaan jalostuvan, kun siitä karsitaan pois sen vähiten merkitsevä osa. Havaitun aineiston oletetaan muodostuvan *piilevien* (engl. ”latent”) muuttujien yhteisvaikutuksista (Bingham 2003). Dimension pienennyksen tarkoituksena on tuoda esiin tärkeimmät näistä piilevistä muuttujista. Se voidaan tulkita myös yrityksenä vähentää aineiston kohinaa.

Vektoriavaruusmallissa kadotetaan yleensä sanajärjestys ja lauseiden järjestys (Bingham 2003). Muun muassa negaatiot ja attribootit menettävät tällöin merkityksensä. Sen vuoksi vektoriavaruusmallilla ei voida tutkia aineiston tarkkaa sisältöä, vaan se soveltuu paremmin esimerkiksi puheenaiheiden selvittämiseen ja vertailuun. Mallilla on mahdollista tutkia myös esimerkiksi sanaparien esiintymistä, jolloin osa sanojen välisistä yhteyksistä säilytetään, mutta tällöin avaruuden ulottuvuuksien määrä kasvaa moninkertaiseksi ja aineistosta tulee harvempi, mikä hidastaa sen käsittelyä ja tekee näytteiden vertailemisesta tilastollisessa mielessä epätarkempaa.

Mallin toiminta perustuu siihen, että eri näytteistä löytyy samoja sanoja.

Samankaltaisuuden vertailu on helpointa toteuttaa vertailemalla sanoja toisiinsa kirjain kerrallaan. Jos kahden sanan kirjaimet ovat täsmälleen samat, myös sanat ovat samat.

Ongelmaksi muodostuu kuitenkin helposti aineiston harvuus, koska luonnollisissa kielissä, erityisesti suomen kielessä, sanat voivat olla useissa eri taivutusmuodoissa.

Näytteiden välisten yhteyksien määrän kasvattamiseksi aineistoa on yleensä esikäsiteltävä. Sanan taivutusmuoto pitää sisällään jonkin verran informaatiota, mutta jättämällä tämä huomiotta saadaan lähes samaa tarkoittavat sanat yhdistettyä yhdeksi sana-avaruuden ulottuvuudeksi. Yhdistely voidaan tehdä esimerkiksi muuttamalla kaikki sanat perusmuotoonsa tai katkaisemalla sanojen päätteet siten, että käytetään vain sanojen vartaloita.

## 2.4.2 Dimension pienennys

Avaruuden dimension eli sen ulottuvuuksien määrän pienennys voidaan tehdä monilla

eri menetelmillä. Yleisessä tapauksessa dimensiota pienennettäessä hukataan osa avaruuden informaatiosta. Aina näin ei kuitenkaan ole, esimerkiksi jos lähtöavaruuden ulottuvuuksien informaatio sisältää redundanssia. Tällöin sama informaatio voidaan esittää myös pienemmällä määrällä ulottuvuuksia. Myös niissä tapauksissa, kun informaatiota joudutaan hävittämään, voidaan hävitettäväksi valita informaation kaikkein vähiten merkitsevä osa.

Eri dimensionpienennysmenetelmät tekevät erilaisia tulkintoja informaation merkitsevyydestä, ja sallivat siten aineistolle erityyppistä vääristymistä. Tätä vääristymistä ei pidä kuitenkaan ehdottomasti minimoida, vaan kuten edellisessä kappaleessa todettiin, aineistoa halutaan nimenomaan jalostaa poistamalla siitä kohinaa. Yleisesti käytettyjä dimensionpienennysmenetelmiä ovat muun muassa *pääkomponenttianalyysi (PCA)*, *riippumattomien komponenttien analyysi (ICA)* sekä *itseorganisoituva kartta (SOM)*.

#### **2.4.2.1 Pääkomponenttianalyysi**

Pääkomponenttianalyysi (engl. ”Principal Component Analysis”, lyh. PCA) on Pearsonin vuonna 1901 tai Hotellingin vuonna 1933 kehittämä dimensionpienennysmenetelmä, tosin eri lähteissä menetelmän kehittäjiksi on nimetty lukuisia muitakin henkilöitä. Pääkomponenttianalyysi pienentää aineiston, jonka muuttujat ovat toisistaan riippuvia, dimensiota säilyttäen mahdollisimman suuren osan aineiston varianssista. Analyysi antaa tuloksena toisen avaruuden, jonka kantavektorit ovat lähtöavaruuden kantavektorien lineaarikombinaatioita, niin sanottuja *pääkomponentteja*. Avaruuden muodostumisen kriteerinä on, että ensimmäinen pääkomponentti on aineistosta löytyvän suurimman varianssin suuntainen. Kyseinen kantavektori asettuu siis kuvaamaan aineiston suurinta vaihtelua. Jos aineisto projisoidaan ensimmäisen pääkomponentin suuntaisesti aliavaruuteen, toinen pääkomponentti on saadun projektion suurimman varianssin suuntainen, ja niin edelleen. Toisin sanoen, myöhemmät pääkomponentit kuvaavat aineiston jäljellä olevan informaation suurinta vaihtelua, jota aiemmat pääkomponentit eivät ole vielä kuvanneet. Ehdosta seuraa, että kaikki pääkomponentit ovat toisiaan vastaan kohtisuorassa. (Jolliffe

2002)

Toimenpide voidaan selittää myös siten, että aineisto pyöritetään avaruudessa sellaiseen asentoon, jossa kukin ulottuvuus esittää aineiston varianssista niin suuren osan kuin yhdessä ulottuvuudessa on mahdollista esittää ilman, että se toistaa aiempien ulottuvuuksien esittämää informaatiota. Pääkomponenttianalyysi järjestää avaruuden siten, että voimakkaasti keskenään korreloivat ulottuvuudet saavat yhteisen merkityksen, jota yksittäinen pääkomponentti kuvaa.

Pelkkä aineiston pyöritys ei kuitenkaan muuta sen sisältämää informaatiota, eikä dimensiota. Pääkomponenttianalyysi antaa tuloksena dimensioltaan yhtä suuren avaruuden, kuin mitä käytetyssä aineistossa on ulottuvuuksia. Ensimmäiset pääkomponentit erottelevat tehokkaimmin aineiston voimakkainta vaihtelua. Myöhemmät pääkomponentit erottelevat yksityiskohtaisemmin informaation jäljelle jäävää osaa. Loppupään pääkomponentit sisältävät kohinaa eli merkityksetöntä satunnaisvaihtelu. Tämän vuoksi ulottuvuuksia karsitaan pois alkaen loppupäästä.

Pääkomponenttianalyysin tuloksessa huomioitavien pääkomponenttien lukumäärä vaikuttaa lopputulokseen. Ei kuitenkaan ole olemassa yksittäistä oikeaa tapaa määrittää säilytettävien pääkomponenttien lukumäärää, vaan sopiva määrä täytyy selvittää tapauskohtaisesti.

#### **2.4.2.2 Riippumattomien komponenttien analyysi**

Riippumattomien komponenttien analyysi (engl. ”Independent Component Analysis”, lyh. ICA) on Pierre Comon:n vuonna 1992 kehittämä toinen yleinen aineiston piilevien rakenteiden löytämiseen käytetty menetelmä (Bingham 2003). Sen tulosavaruuden ulottuvuuksia kutsutaan *riippumattomiksi komponenteiksi*, ja nimensä mukaisesti ne ovat toisistaan tilastollisesti riippumattomia. Samoin kuin pääkomponenttianalyysin tuottamat pääkomponentit, myös riippumattomat komponentit ovat aineistona käytettävän avaruuden ulottuvuuksien lineaarikombinaatioita. Toisin kuin pääkomponenttien, riippumattomien komponenttien ei vaadita olevan keskenään ortogonaalisia. Tilastollisen riippumattomuuden vaatimus ei ole absoluuttisesti heikompi ehto kuin

ortogonaalisuus, koska kaksi keskenään ortogonaalista ulottuvuutta voivat kuitenkin olla toisistaan tilastollisesti riippuvia, esimerkiksi aineiston sijoittuessa paraabelin muotoiselle käyrälle. (Comon 1992)

Riippumattomien komponenttien analyysia on perinteisesti käytetty lähinnä signaalin käsittelyyn. Se soveltuu erityisen hyvin niin sanotun *cocktailkutsuilmion* (engl. ”cocktail party effect” tai ”blind source separation”) ratkaisuun. Ilmiössä on kyse eri informaatiolähteiden erottelusta yksittäisen signaalin perusteella. Riippumattomien komponenttien analyysin käyttö tekstianalyysissa on soveltavaa. Tulokset ovat kuitenkin osoittautuneet hyviksi. (Bingham 2003)

Riippumattomien komponenttien analyysin tulosta voidaan parantaa esikäsittelemällä aineisto esimerkiksi pääkomponenttianalyysilla, singulaariarvohajotelmalla (engl. ”Singular Value Decomposition”, lyh. SVD) tai satunnaisprojektiolla (engl. random projection). Esikäsittelyllä voidaan parantaa analyysin tulosta, koska riippumattomien komponenttien analyysilla on taipumusta ylioppimiseen etenkin tapauksissa, joissa aineistoa on vähän suhteessa käsiteltävän avaruuden dimensioon. (Bingham 2003)

#### **2.4.2.3 Itseorganisoituva kartta**

Teuvo Kohonen kehitti vuonna 1982 aineiston ryhmittelyyn ja visualisointiin tarkoitetun neuraalilaskentaan perustuvan menetelmän nimeltään *itseorganisoituva kartta* (engl. ”Self-Organizing Map” tai ”Kohonen map”, lyh. SOM), joka esittää aineiston kaksi- tai useampiulotteisella solujen verkostosta muodostuvalla kartalla. Menetelmä sijoittaa toisiaan muistuttavat näytteet lähellä toisiaan oleviin kartan soluihin. Aineiston klusterirakenne voidaan visualisoida esimerkiksi värittämällä erityyppisiä näytteitä sisältävät solut eri väreillä. (Kohonen 1982) (Lagus *et al.* 1996)

Itseorganisoituva kartta eroaa muista yleisesti käytetyistä menetelmistä siinä, että se ei käsittele aineistoa lineaarisesti. Kaksiulotteisessa tapauksessa se esittää moniulotteiselle pinnalle asettuneen aineiston ikään kuin taivutteleamalla ja silottelemalla aineiston tasoon.

Visualisoinnissa käytettävät kaksi ulottuvuutta pitävät sisällään liian vähän informaatiota monien tekstianalyysiongelmien ratkaisemiseksi, ja menetelmän tulokset useammalla ulottuvuudella ovat vaikeasti tulkittavia. Menetelmän suorituskky on kuitenkin todettu hyväksi tekstikäsitteilynkin tapauksessa. (Bingham 2003)

#### **2.4.2.4 Muut dimensionpienennysmenetelmät**

*Faktorianalyysi* (engl. ”factor analysis”) on psykologiassa ja yhteiskuntatieteissä usein käytetty aineiston piilevien muuttujien löytämiseen tarkoitettu menetelmä. Sen tulokset ovat periaatteessa yhteneviä pääkomponenttianalyysin kanssa, tosin lopputuloksen esitysmuoto on hieman erilainen. (Bingham 2003)

*Kognitiivisten komponenttien analyysi* (engl. ”COgnitive Component Analysis”, lyh. COCA) on ohjaamaton ryhmittelymenetelmä, jonka tulosten on tarkoitus olla linjassa ihmisen kognitiivisen tiedonkäsittelyn kanssa. Menetelmällä ei ole kuitenkaan tällä hetkellä selkeää laskennallista muotoa, koska ihmisen tiedonkäsittelyprosessi ei ole yksityiskohtaisesti selvillä. On kuitenkin löytynyt mielenkiintoisia viitteitä siitä, että hyvin lähelle ihmisen tiedonkäsittelyn tuloksia päästään aiemmin esitellyllä riippumattomien komponenttien analyysillä. Kyseinen havainto on tehty muun muassa tilastollisen tekstianalyysin kohdalla. (Hansen *et al.* 2005) (Feng & Hansen 2006)

Muita lineaarisia menetelmiä dimension pienennykseen ovat muun muassa ”*projection pursuit*”, ”*non-negative matrix factorization*” (lyh. NMF), ”*probabilistic latent semantic analysis*” (lyh. PLSA), ”*latent Dirichlet allocation*” (lyh. LDA), ”*multinomial PCA*” (lyh. MPCA) ja ”*mixture modeling*”, joille ei löydy yleisesti käytettyjä suomennoksia. (Bingham 2003)

*Paikallinen pääkomponenttianalyysi* (engl. ”local PCA”) suoritetaan jakamalla aineisto osiin ja suorittamalla pääkomponenttianalyysi erikseen kullekin osalle, jolloin kyseessä ei ole enää lineaarinen muunnos. (Bingham 2003)



### 2.4.3 Latentti semanttinen indeksointi

Pääkomponenttianalyysin käyttöä tekstianalyysissa kutsutaan *latentiksi semanttiseksi indeksoinniksi* (engl. ”Latent Semantic Indexing”, LSI) (Booker *et al.* 1999), joka on yksi vektoriavaruusmallin ohjaamattomista toteutustavoista.

Tekstianalyysissa pääkomponentit voidaan tulkita sanojen yhdistelminä, joihin jokaiseen kunkin sanan esiintyvyys vaikuttaa omalla painokertoimellaan. Pääkomponenteilla on taipumus asettua kuvaamaan sanaryhmiä, joiden jäsenet esiintyvät usein samoissa yhteyksissä. Tällöin sanaryhmien väliset erot tulevat selvemmin esiin kuin alkuperäisessä sana-avaruudessa.

Toinen merkittävä ominaisuus on sanojen merkitysten yhdenmukaistuminen. Koska pääkomponentteja on vähemmän kuin alkuperäisiä sanaulottuvuuksia, tulosavaruudessa, josta tässä työssä käytetään nimitystä ”*aiheavaruus*”, ei ole yhtä paljon informaatiota kuin sana-avaruudessa. Pääkomponenttianalyysi karsii informaatiota siten, että toisiaan muistuttavat tekstinäytteet sijoittuvat samaan paikkaan avaruudessa. Toisaalta myös yhdessä esiintyneitä sanoja vastaavat ulottuvuudet voidaan esittää uudessa avaruudessa enemmän yhdensuuntaisina, kun alunperin ne olivat kaikki toisiaan vastaan kohtisuorassa. Tällöin näytteen sijainnin kannalta on lähes yhdentekevää, onko näytteessä käytetty jotakin tiettyä sanaa vai sen kanssa usein esiintyvää toista sanaa. Tämä on eduksi silloin, kun halutaan vertailla puheenaiheita eri näytteiden välillä, koska puheenaiheet tunnistetaan samaksi myös silloin, kun sanavalikoima on ollut hieman erilainen.

Pääkomponenttianalyysiin tarvittava muunnosmatriisi voidaan ratkaista joko aineiston singulaariarvohajotelman tai sen kovarianssimatriisin ominaisarvohajotelman avulla. Kovarianssimatriisi  $C$  voidaan laskea aineiston  $F$  käänteismatriisin  $F^{-1}$  ja otoskeskiarvon  $\bar{x}$  avulla kaavalla

$$C = F^{-1} \bar{x} \bar{x}'$$

Kovarianssimatriisin  $C$  ominaisarvohajotelma voidaan esittää muodossa

$$V^{-1}CV = D$$

missä  $D$  on  $C$ :n ominaisarvoista muodostettu diagonaalimatriisi. Ominaisvektorit sisältävä matriisi  $V$  voidaan ratkaista laskennallisesti monien matemaattisten apuohjelmien avulla. (Jolliffe 2002)

$V$ :n kukin sarakevektori  $v$  on yksi kovarianssimatriisin  $C$  ominaisvektoreista. Ominaisvektorit  $v$  ovat aineiston  $F$  pääkomponentteja. Ne järjestetään niitä vastaavien ominaisarvojen suuruuden mukaiseen järjestykseen, jolloin ne ovat myös merkitsevyysjärjestyksessä. Kutakin vektoria  $v_i$  vastaava ominaisarvo löytyy diagonaalimatriisin  $D$  kohdasta  $(i, i)$ . (Jolliffe 2002)

Ennen lineaarimuunnosta aineisto voidaan keskiarvoistaa vähentämällä siitä sen keskiarvon estimaatti. Käytännön tilanteissa estimaattina käytetään koko aineiston keskiarvoa (Bingham 2003). Keskiarvoistamisen oikeutuksesta on vaihtelevia mielipiteitä. Hyvänä puolena on, että keskiarvoistetun aineiston ratkaisu on pienimmän neliösumman mielessä optimaalinen, eli mikä tahansa määrä ensimmäisiä pääkomponentteja kuvaa aineiston tällöin pienimmän neliösumman mielessä parhaalla mahdollisella tavalla säilyttäen mahdollisimman paljon informaatiota. Huonona puolena voidaan nähdä se, että muunnoksessa ei ole keskiarvoistamisen yhteydessä kyse pelkästä avaruuden pyörittämisestä, vaan myös siirrosta. Tulosavaruuden ulottuvuudet ovat siten vielä abstraktimpia ja vaikeammin ymmärrettäviä alkuperäisen aineiston kannalta.

Sovelluskohteesta kuitenkin riippuu, onko keskiarvoistaminen perusteltua vai ei (Miranda *et al.* 2008). Tekstianalyysin tapauksessa keskiarvoistus voidaan tehdä, koska analyysin lähtökohtana toimivat sanaulottuvuudet ovat jo itsessään niin abstrakteja, ettei niitä yritetä sinänsä tulkita. Etuna saadaan helpommin ymmärrettävä tulosavaruus, jossa aineisto on tasaisesti jakautunut keskipisteen ympärille.

### **3 Kehityskohde**

Diplomityön kokeellisessa osuudessa toteutettiin palvelu, jolla pyritään helpottamaan kaupunkisuunnittelussa hyödynnettävän paikallistietämyksen käsittelyä sekä parantamaan paikallisten toimijoiden osallistamista visualisoimalla heidän antamien tekstimuotoisten kommenttien puheenaiheita paikkatietojärjestelmän kartalla. Palvelu toteutettiin Aalto-yliopiston teknillisen korkeakoulun Arkkitehtuurin laitoksen Yhdyskunta- ja kaupunkisuunnittelun yksikön tuottaman ja ylläpitämän Espoon Keskusten Kehittämisfoorumin paikkatietojärjestelmän osana.

#### **3.1 Kehittämisfoorumi**

*Kehittämisfoorumi* on Internet-sivusto, joka tarjoaa yksittäisen alueen asukkaille tietoa alueen tapahtumista ja suunnitteilla olevista hankkeista, sekä mahdollisuuden ottaa kantaa asuinalueensa kehitykseen (Espoon Keskusten kehittämisfoorumin ylläpito 2010). Kommentointimahdollisuus on toteutettu paikkatietojärjestelmään siten, että käyttäjä voi sijoittaa valitsemiinsa kartan pisteisiin tekstimuotoisia kommentteja. Kommenttien tarkoituksena on antaa kaupunkisuunnittelijoille tietoa alueen asukkaiden ja muiden osallisten yleisistä mielipiteistä, sekä suhtautumisesta yksittäisiin asioihin, kuten tiettyihin maastonkohtiin ehdotettuihin muutoksiin.

##### **3.1.1 Kehittämisfoorumin käyttö kaupunkisuunnittelussa**

Sivuston käyttäjiä ovat paikalliset toimijat, järjestelmän ylläpitäjät ja kaupunkisuunnittelijat. Sivusto on kaikille avoin ja sitä voi käyttää milloin haluaa. Sisällöllinen ylläpito tapahtuu Arkkitehtuurin laitoksella. Kommentit käydään ylläpidon toimesta läpi päivittäin tai tarpeen mukaan, ja ne julkaistaan nähtäväksi paikkatietojärjestelmän karttanäkymässä. (Ylläpidon ohje kerätyn materiaalin analysoinnista 2009) (Espoon Keskusten kehittämisfoorumin ylläpito 2010)

Ylläpidon tehtäviin kuuluu myös raportointi Espoon kaupungin suunnittelijoille kerran puolessa vuodessa. Raportti muodostetaan käymällä läpi kommentteja ja etsimällä niistä vahvimmin esille nousevia näkemyksiä. Aineisto luetaan läpi moneen kertaan, ja niistä

pyritään muodostamaan hyvä kokonaiskuva. Eräänä menetelmänä ylläpitäjät käyttävät kommenttien jakamista eri teemoihin. Raportin lisäksi kaikki kommentit lähetetään kaupunkisuunnittelijoille. Kaupunkisuunnittelijat voivat myös itse lukea palveluun jätettyjä yksittäisiä kommentteja samalla käyttöliittymällä, jota kansalaiset ja ylläpitäjät käyttävät. (Ylläpidon ohje kerätyn materiaalin analysoinnista 2009)

Ylläpitäjien mukaan raporttia luotaessa kokonaiskuvan muodostaminen on työlästä. Vaikeaa on havaita etenkin, minkä asioiden takana on isompi kannatus ja mitkä palautteista taas ovat vähäpätöisempiä mielenilmauksia. Tarvitaan työkalu, joka helpottaa kaupunkisuunnittelijoiden ja ylläpidon työtä ja auttaa muodostamaan objektiivisemmän kuvan osallisten mielipiteistä.

### **3.1.2 Aineisto**

Kehittämisfoorumille jätetyt kommentit sisältävät koordinaattitiedon lisäksi muita määrämuotoisia tietoja sekä vapaan tekstisisällön. Kommenttia luodessaan käyttäjä valitsee kommentin tyypin kolmesta vaihtoehdosta: ”positiivinen”, ”negatiivinen” tai ”kehittämisehdotus”. Kommentille kirjoitetaan otsikko ja sisältö tekstimuodossa. Lisäksi kommentin metatietoihin tallentuu automaattisesti kommentoinnin ajanhetki ja käyttäjän aiemmin syöttämät ikä, sukupuoli sekä tieto siitä, asuuko kommentoija itse kehitettävällä alueella.

Espoon Keskuksen kehittämisfoorumilla on työn tekohetkellä 208 suomenkielistä kommenttia.

## **3.2 Järjestelmän teknologiat**

Palvelun kehityksen aikana Kehittämisfoorumi siirrettiin uudelle tekniselle alustalle ja samassa yhteydessä siihen tehtiin suuria käyttöliittymämuutoksia. Palvelu toteutettiin toimimaan uudella alustalla.

### 3.2.1 Sisällönhallintajärjestelmä

*Sisällönhallintajärjestelmä* on järjestelmä, joka mahdollistaa monimutkaisen verkkosivuston luomisen ja hallinnoimisen suhteellisen helposti. Eri järjestelmät tarjoavat erilaisia ominaisuuksia, joita verkkosivustolle voi luoda. Lisäksi järjestelmien käytettävyys eroaa toisistaan huomattavasti.

Kehittämisfoorumin sisällönhallintajärjestelmänä oli aiemmin ollut *Joomla!* ja se siirrettiin toimivaksi *Drupal*:lla. Oleellisinä erona kehittämisfoorumin kannalta on se, että Joomla! ei voi hallinnoida kartalla olevia kommentteja suoraan omina elementteinään, ja Drupal taas voi. Tämä helpottaa huomattavasti lisäominaisuuksien luomista, koska karttakommentteihin voidaan viitata suoraan Drupalin tarjoamista muista palveluista, kuten keskustelupalstalta. Myös Drupalin käytettävyys on todettu Yhdyskunta- ja kaupunkisuunnittelun yksikön ylläpitämän toisen järjestelmän, Koulufoorumin, käytössä paremmaksi kuin Joomla!:n.

Sekä Drupalin että Joomla!:n ominaisuuksia voidaan laajentaa *php*-ohjelmointikielellä, joka on verkkosivustoja kuvaavan html-kielen kaltainen, mutta mahdollistaa toiminnallisen sisällön luomisen.

### 3.2.2 Karttapalvelu

Interaktiivisen kartan luomiseksi verkkosivustolle tarvitaan *karttapalvelu*, joka päivittää karttakuvan vastaamaan tietokannassa olevaa paikkatietoa. Kuten sisällönhallintajärjestelmät, myös karttapalvelut eroavat toisistaan ominaisuuksiltaan ja käytettävyydeltään.

Vanhassa alustassa Joomla! ja *Mapserver*-karttapalvelu oli yhdistetty siten, että kartalle piirrettyyn pisteeseen voitiin linkittää Joomla!:ssa luotu artikkeli. Tällöin Joomla!:lla voitiin hallinnoida karttanäkymää.

Uudessa järjestelmässä Mapserverin korvaajaksi valittiin *OpenLayers*, koska sen käytettävyys sekä järjestelmän kehittäjän että loppukäyttäjän kannalta on todettu

paremmaksi. Esimerkiksi Mapserverin karttanäkymän vierittäminen tapahtuu suurina nykyisinä klikkaamalla kartan reunalla olevia nuolia, kun taas OpenLayersissä näkymää voi vierittää vapaasti painamalla hiiren napin pohjaan ja liikuttamalla hiirtä. OpenLayers on käytössä huomattavasti nopeampi. Lisäksi siinä voidaan näyttää *Google Mapsin* tarjoamaa ilmaista karttanäkymää, mikä oli etenkin järjestelmän siirtymävaiheen aikana hyödyllinen ominaisuus, kun Espoon Kaupungin tarjoamaa pohjakarttaa ei oltu vielä siirretty OpenLayersin käyttämään koordinaatistoon.

OpenLayers-karttapalvelun lisäämiseksi verkkosivustolle tarvitaan *JavaScript*-ohjelmointikieltä, joka vastaa interaktiivisen karttanäkymän ajamisesta käyttäjän omalla tietokoneella.

### 3.2.3 Palvelin

Palvelimen tarkoituksena on ylläpitää sisällönhallintajärjestelmää suorittamalla sen antamat komennot sekä pitää yhteyttä käyttäjiin ja tietokantaan. Järjestelmän palvelimena toimi ennen siirtoa ja sen jälkeen *Apache*.

### 3.2.4 Tietokanta

Tietokannassa säilytetään suuria tietomääriä ja sitä hallinnoidaan *tietokannan hallintajärjestelmällä*. Tietokannan hallintajärjestelmä tarjoaa sekä ohjelmallisen että visuaalisen käyttöliittymän, joilla kannassa olevaa tietoa voi käsitellä. *MySQL* mahdollistaa tietokannan käsittelyn php-ohjelmakoodista käsin, sekä erillisen verkkosivuston, *MyAdmin* -hallintapaneelin, kautta.

Kehittämisfoorumilla käytössä olevassa *MySQL 5.0*:ssa on tuki paikkatiedolle. Se tarjoaa erityisesti paikkatiedon käsittelyä helpottavia toimenpiteitä, kuten pisteiden ja monikulmioiden tallentamisen sekä etäisyyksien laskennan. Kehittämisfoorumin kommenttien koordinaatit on tallennettu paikkatietomuodossa erilliseen geometriatauluun.

Alustan vaihdon yhteydessä kommenttien tekstisisältöjen muoto tietokannassa

vaihdettiin samalla toisenlaiseksi.

### **3.3 Palvelun vaatimukset**

Suunniteltavan palvelun tavoitteena on parantaa osallisten jättämien kommenttien käsittelyprosessia tekemällä siitä nykyistä käsittelyä tehokkaampi, helppokäyttöisempi, oikeudenmukaisempi, läpinäkyvämpi ja houkuttelevampi.

Ennen uuden palvelun suunnittelua ja toteutusta on kartoitettava mitä sen halutaan tekevän ja millä ehdoilla. Ohjelmistokehityksen yleiset periaatteet, eri käyttäjäryhmät sekä muutoksen tarkoitus aiheuttavat palvelulle ja sen kehittämisprosessille erilaisia vaatimuksia. Myös ongelman ratkaisemiseksi valittu menetelmä voi aiheuttaa itsessään tiettyjä vaatimuksia, mutta ne voidaan kartoittaa vasta kun menetelmä on valittu. Sen vuoksi ne on jätetty tästä kappaleesta pois ja esitetty myöhemmin kappaleessa 4.4 menetelmän esittelyn yhteydessä.

Vaatimusten tarkoituksena on auttaa kehitystyön lopputuloksen arvioimisessa sekä suunnata työn kulkua oikeaan suuntaan. Yksittäisten vaatimusten toteutumattomuus ei ole este palvelun käyttöönotolle, mutta kunkin vaatimuksen täyttymistä on tavoiteltava hyödyllisiä tuloksia antavaan lopputulokseen pääsemiseksi.

Seuraavissa kappaleissa on käsitelty eri lähteiden palvelulle asettamia vaatimuksia. Kunkin kappaleen loppuun on taulukoitu kappaleessa esitetyt vaatimukset vaatimuslähteelle annetun tunnuskirjaimen sekä juoksevan numeron mukaan.

#### **3.3.1 Käyttäjien tarpeet**

Palvelun ensisijaisia käyttäjiä ovat kaupunkisuunnittelijat, koska heidän käsityksellään kommenttien sisällöstä on suora vaikutus päätöksentekoon, ja he voivat aiheuttaa siten todellista muutosta suunniteltavassa ympäristössä. Toissijaisia käyttäjiä ovat osalliset sekä järjestelmän ylläpitäjät ja kehittäjät. Palvelun ominaisuuksia päätettäessä keskitytään ensisijaisesti kaupunkisuunnittelijoiden tarpeisiin, jotka tosin ovat suurilta osin päällekkäisiä muiden käyttäjäryhmien tarpeiden kanssa. Esimerkiksi läpinäkyvyys

ja houkuttelevuus eivät ole kaupunkisuunnittelijoiden kannalta oleellisia ominaisuuksia, vaan niiden tarkoituksena on antaa ominaisuudesta parempi kuva julkisuudessa ja kerätä palvelulle lisää käyttäjiä ja huomioarvoa. Siksi ominaisuuden kehityksessä panostetaan enemmän tehokkuuteen, helppokäyttöisyyteen sekä kommenttien tasapuoliseen käsittelyyn.

Kaupunkisuunnittelijat ja Kehittämisfoorumin ylläpitäjät ovat peräänkuuluttaneet apuvälinettä, jolla kommenttikartan kommentteista saisi helpommin muodostettua kokonaiskuvan. Jotta palvelu vähentäisi kaupunkisuunnittelijoiden työmäärää eikä lisäisi sitä, sen on oltava riittävän tehokas käytettävyydeltään. Helppokäyttöisyys myös alentaa etenkin uusien käyttäjien kohdalla kynnystä palvelun käyttämiseen.

Analyysin tulokset on esitettävä visuaalisesti, jotta suuri informaatiomäärä saadaan välitettyä käyttäjälle mahdollisimman tehokkaasti. Kuten kappaleessa 2.3 todettiin, visuaalinen esitystapa sopii erityisen hyvin tilanteisiin, joissa eri käyttäjäryhmillä on erilaiset valmiudet tulkita abstraktia informaatiota.

Palvelun ei toivota analysoivan yksittäisiä kommentteja, vaan kokonaisuutta, koska nimenomaan kokonaisuuden hahmottaminen on koettu vaativaksi manuaalisessa käsittelyssä. Yksittäisten kommenttien sisällön analysointi automaattisella tekstianalyysillä olisi myös erittäin haasteellista kielen monimutkaisuudesta johtuen.

Palvelun ei haluta rajoittavan suunnittelijoiden toimintaa, vaan sen halutaan tuovan lisämahdollisuuksia kommenttien käsittelyyn. Suunnittelijoiden on tarvittaessa voitava käydä läpi kaikki kommentit kuten ennenkin, jotta kaikki tieto on edelleen käytettävissä.

Analyysin tuloksen on paljastettava jotakin todellista informaatiota siten, että sen perusteella voidaan tehdä johtopäätöksiä. Yhdyskunta- ja kaupunkisuunnittelun yksikössä on aiemmin käytetty itseorganisoituvaa karttaa muunlaisen aineiston käsittelyyn ja kokemus oli, että käyttäjä ajautuu helposti näkemään analyysin tuloksessa sellaisia hahmoja kuin itse olettaa tai toivoo näkevänsä. Tulokset olivat toisin sanoen liian positiivisia. Sen sijaan että ne olisivat tuoneet käyttäjälle uutta informaatiota, ne vain vahvistivat niitä käsityksiä, joita käyttäjällä ennestään oli, riippumatta käsityksen



oikeellisuudesta. Tätä näkökohtaa on arvioitava kriittisesti palvelua kehitettäessä sekä sen toiminnan tuloksia arvioitaessa.

Palvelun on otettava huomioon kommenttien spatiaalinen luonne. Kommentit on sijoitettu kartalle nimenomaan kuvaamaan paikkoja tai alueita ja tätä tietoa halutaan hyödyntää analyysissa. Kommenttien puheenaiheet ja sijainnit halutaan yhdistää informatiivisella tavalla.

*Taulukko 1, Palvelun vaatimukset käyttäjien kannalta K1 – K7*

Vaatus K1	Informatiivisuus
Vaatus K2	Tiedon oikeellisuus
Vaatus K3	Visuaalinen esitystapa
Vaatus K4	Aineiston spatiaalisuuden huomioon ottaminen
Vaatus K5	Aineiston kokonaisvaltainen käsittely
Vaatus K6	Ei rajoita muuta käyttöä
Vaatus K7	Helppokäyttöisyys

### **3.3.2 Palvelun sopivuus olemassa olevaan järjestelmään**

Koska palvelu toteutetaan osana laajempaa järjestelmää, sen on vastattava järjestelmän suunnitteluperiaatteita ja standardeja. Kehittämisfoorumin sivuston esittelyssä luonnehditaan sivustoa *avoimeksi, ajankohtaiseksi, puolueettomaksi ja paikalliseksi*. Lisäksi foorumin sanotaan kokoavan, yhdistelevän ja esittävän tietoa helppokäyttöisellä tavalla, mikä on yhtenevää edellisen kappaleen informatiivisuus-, visuaalisuus ja helppokäyttöisyysvaatimusten K1, K3 ja K7 kanssa. Ylläpidon tulee olla edullista ja

paikallisesti toteutettua. Palvelun tavoitteena on tukea alueen monipuolista kehittämistä ja imagon parantamista. (Espoon Keskusten kehittämisfoorumin ylläpito 2010)

Kehittämisympäristön ei ole tarkoitus olla ainutlaatuinen, vaan vastaavia ympäristöjä on kehitteillä useille eri alueille. Tämä on otettava huomioon myös palvelun toteutuksessa, eli sen on oltava helposti kopioitavissa muihin vastaaviin ympäristöihin.

Palvelun käyttöliittymän on oltava riittävän selkeä, eikä se saa viedä liikaa huomiota järjestelmän muilta ominaisuuksilta. Työn aiheena ei ole käytettävyyden kehittäminen, joten työn puitteissa käytettävyyksvaatimukseksi riittää yhdenmukaisuus järjestelmän muiden osien kanssa. Vaatimus on osittain päällekkäinen helppokäyttöisyysvaatimuksen K7 kanssa.

*Taulukko 2, Palvelun yhteensopivuusvaatimukset Y1 – Y8*

Vaatimus Y1	Alueen kehityksen tukeminen
Vaatimus Y2	Avoimuus ja läpinäkyvyys
Vaatimus Y3	Puolueettomuus
Vaatimus Y4	Ajankohtaisuus
Vaatimus Y5	Ylläpidon edullisuus ja paikallisuus
Vaatimus Y6	Alueen imagon parantaminen
Vaatimus Y7	Käytettävyydeltään yhdenmukainen
Vaatimus Y8	Kopioitavuus

### **3.3.3 Vaatimukset osallistavuuden parantamiseksi**

Osallistavuuden perusvaatimuksia ovat tasavertaisuus ja puolueettomuus (Dragičević &

Balram 2004). Vaikka sivuston esittelyssä (Espoon Keskusten kehittämisfoorumin ylläpito 2010) mainitaan sivuston puolueettomuus, tällä ei kuitenkaan tarkoiteta ehdotonta tasavertaisuutta. Liian voimakas tasapuolisuusvaatimus asettaa tiukkoja rajoja palvelun ja ylläpidon toiminnalle ja voi jopa vaikeuttaa kokonaiskuvan muodostamista. Esimerkiksi paljon kirjoitusvirheitä sisältävää tai muulla tavalla tavallisesta poikkeavaa tekstiä on vaikea analysoida automaattisesti, koska yhteyksiä muihin teksteihin ei löydy. Vastaavasti keskeisiä teemoja käsitteleviä ytimekkäitä tekstejä on joskus hyödyllistä käyttää esimerkinomaisesti edustamaan kommenttiryhmän sisältöä, jolloin valitut kommentit tulevat muita enemmän esille. Siksi riittää, että kommentteja käsitellään lähtökohtaisesti tasapuolisesti, vaikka osa niistä lopulta saisikin suuremman huomioarvon kuin toiset.

Kommenttikartta ei ole osa kaupungin virallista palautejärjestelmää (Espoon Keskusten kehittämisfoorumin ylläpito 2010). Sen vuoksi ei ole myöskään juridisia vaatimuksia sille, että järjestelmän pitäisi täyttää tietyt tasavertaisuusehdot. Tasavertaisuus kuitenkin parantaa oleellisesti osallistavuutta ja siten kansalaisten mahdollisuuksia vaikuttaa asioihin. Sitä on palvelun luonteen mukaisesti tavoiteltava ja sen vuoksi tasavertaisuuden tavoittelu voidaan listata yhtenä palvelun tavoitteista.

Toinen osallistavuuden perusvaatimus, saavutettavuus, liittyy myös aiemmin mainittuun helppokäyttöisyyteen. Saavutettavuudella ei tarkoiteta pelkästään sitä, että käyttäjä pääsee käsiksi järjestelmään, vaan myös että hän kykenee tekemään sillä haluamansa toimenpiteet. Palvelun on oltava riittävän helppokäyttöinen, jotta asiaan perehtymätönkin henkilö voi sitä käyttää.

*Taulukko 3, Palvelun vaatimukset osallistavuuden edistämiseksi O1 ja O2*

Vaatimus O1	Tasavertaisuus
Vaatimus O2	Saavutettavuus

### 3.3.4 Tekniset vaatimukset

Ohjelmiston kehityksessä usein tehtävä virhe on olettaa, että kaikki muutos on kehitystä. Ohjelman jokainen ominaisuus varaa käyttäjän kognitiivisia resursseja, jolloin yhden uuden ominaisuuden lisääminen vähentää muiden ominaisuuksien suhteellista huomioarvoa. Sen vuoksi muutoksen on oltava käyttäjälle mahdollisimman vaivaton. Usein parhaisiin tuloksiin päästään tekemällä ominaisuuksista niin yksinkertaisia ja riisuttuja kuin mahdollista, säilyttäen kuitenkin kaikki oleelliset toiminnot.

Lisäksi ohjelmamuutoksilla on taipumusta kasvattaa ohjelman ominaisuuksien määrän ohella myös sen ylläpidon työmäärää. Jokaisella ominaisuudella on mahdollisuus olla toimimatta sellaisessa ympäristössä, jossa sitä ei ole vielä testattu. Koska koko järjestelmän on tarkoitus säilyä edullisesti ja paikallisesti ylläpidettävänä, on lisäpalvelun ylläpidon vaativuuteen myös kiinnitettävä huomiota. Tämä kuitenkin listattiin jo vaatimukseksi Y5 kappaleessa 3.3.2.

Ohjelmiston kehityksessä ei pidä myöskään sortua liialliseen optimismiin. Kaikesta panostuksesta huolimatta muutos voi osoittautua virheeksi, jolloin järjestelmän kokonaistoiminta ilman sitä on parempaa kuin sen kanssa. Tällöin on eduksi, jos ominaisuuden voi helposti kytkeä pois päältä ilman, että koko järjestelmän lähdekoodia tarvitsee käydä läpi ominaisuuden liittymäkohtien kitkemiseksi. Kannattaa suosia modulaarista lähestymistapaa, jossa yksittäiset ominaisuudet nähdään sekä irrallisina ominaisuuksina että osana kokonaisuutta. Sama koskee myös ominaisuuden dokumentaatiota, esimerkiksi käyttöohjetta. Kun ohjeesta tehdään erillinen dokumentti, sen voi tarvittaessa helpommin poistaa.

Lisäksi erityisesti ohjelmistotuotannossa voi helposti aliarvioida tarvittavan työmäärän. Ohjelmamuutoksen tekemiseksi ei riitä pelkkä lähdekoodin kirjoittaminen, vaan aikaa on varattava myös dokumentointiin, yksikkö-, integrointi- ja systeemitestaukseen, käyttöönottoon sekä häiriötilanteisiin.

Vuonna 2004 Fraunhofer institute AIS:ssä tehdyssä paikkatietojärjestelmän kehitykseen liittyvässä kokeessa (Voss *et al.* 2004) listattiin kahden järjestelmän osan yhteen

liittämisen teknisiksi vaatimuksiksi osien *maksimaalinen itsenäisyys*, *nopea vasteaika*, *aineiston synkronointi*, *redundantin informaation välttäminen* sekä *verkon palomuurien huomioon ottaminen*. Näitä vaatimuksia voidaan soveltaa myös analyysipalvelun kehittämisessä ja integroimisessa osaksi kehittämisfoorumia. Järjestelmän osien maksimaalisen itsenäisyyden vaatimus on yhtenevä modulaarisuusvaatimuksen kanssa. Nopea vasteaika on oleellinen tekninen vaatimus, joka parantaa palvelun käytön tehokkuutta. Aineiston synkronoinnilla tarkoitetaan sitä, että aineiston muutokset päivittyvät automaattisesti kaikkiin järjestelmän osiin. Toisin sanoen, analyysin on aina käsiteltävä paikkatietojärjestelmässä kyseisellä hetkellä olevaa, eikä vanhentunutta, aineistoa. Redundantin informaation välttäminen liittyy samaan asiaan. Redundanssia syntyy, kun samaa tietoa säilytetään useammassa eri paikassa. Tämä ei yleensä ole ongelma muulloin kuin aineiston synkronoinnin ongelmatapauksissa tai muistitilan ollessa rajallinen. Palomuurien huomioonottaminen liittyi kyseisessä integraatiossa järjestelmän erityistarpeisiin verkkoliikenteen suhteen, joten sitä ei voida suoraan soveltaa tähän työhön. Vaatimus voidaan kuitenkin yleistää palvelun alustariippumattomuudeksi, jolla tarkoitetaan sitä, että palvelua pitää voida käyttää erilaisilla pääteohjelmilla ja -kokoonpanoilla.

*Taulukko 4, Palvelun tekniset vaatimukset T1 - T6*

Vaatimus T1	Muutoksen yksinkertaisuus käyttäjän kannalta
Vaatimus T2	Modulaarisuus
Vaatimus T3	Toteuttamiskelpoisuus resurssien puitteissa
Vaatimus T4	Nopea vasteaika
Vaatimus T5	Aineiston synkronointi
Vaatimus T6	Alustariippumattomuus

### **3.3.5 Julkisen interaktiivisen verkkopalvelun menestystekijät**

Laajassa 190:a eurooppalaista kaupunkia koskevassa EDC Survey 1998 -tutkimuksessa kartoitettiin kaupunkien tarjoamien julkisten interaktiivisten verkkopalvelujen menestystekijöitä, joita voidaan soveltaen käyttää palvelun kehittämisen lähtökohtina. (Mino 2000)

Kartoituksen mukaan hyvin menestyneet palvelut on toteutettu osana muita julkisia palveluja. Ne nivoutuvat yhteen muiden palvelujen kanssa saaden niistä synergiaetua.

Ne on toteutettu iteratiivisesti siten, että palvelun toiminnasta on voitu kerätä kokemuksia kesken sen kehittämisen.

Menestyneet palvelut on nähty osana laajempaa teknologista kehitystä. Niillä on ollut monia rinnakkaisia tavoitteita, kuten teknologian kehittäminen, osallistavuuden parantaminen ja tehokkuus.

Palveluilla on julkisen hallinnon tuki, eikä niillä yritetä ohittaa päättäjien toimivaltaa.

On selkeästi määritelty kuka on palvelun kehittäjä, tuottaja ja ylläpitäjä. Kehitys tapahtuu instituutiossa, jolla on kyky toteuttaa tarvittavat tekniset toimenpiteet. Instituutio on osittain itsenäinen julkisesta hallinnosta, jotta se voi vapaasti hyödyntää yksityistä sektoria ja tarvittavia resursseja ilman poliittista painostusta ja byrokraattisia rajoitteita.

*Taulukko 5, Palvelun menestymisen vaatimukset M1 - M6*

Vaatimus M1	Toimivaltainen kehitysyhteisö
Vaatimus M2	Julkisen hallinnon tuki
Vaatimus M3	Roolien selkeys
Vaatimus M4	Osa laajempaa kokonaisuutta
Vaatimus M5	Toteutuksen iteratiivisuus
Vaatimus M6	Monitavoitteisuus

## 4 Kehitettävän palvelun määrittely

Tämän diplomityön osana suunniteltiin interaktiivisia verkkopalveluita, joiden tarkoituksena on helpottaa kaupunkisuunnitteluun liittyvän paikallistiedon käsittelyä, sekä edistää paikallisten toimijoiden osallistamista mukaan kaupunkisuunnitteluprosessiin. Kehitystyön alustaksi valittiin Espoon Keskusten Kehittämisfoorumi. Suunnitelluista palveluista valittiin työn puitteissa toteutettavaksi yksi, osoitukseksi siitä, että tämän tyyppisiä palveluita voidaan kehittää ja että asetetut tavoitteet voidaan saavuttaa. Toteutetun palvelun toimintaa arvioidaan yksityiskohtaisesti aiemmin asetettujen vaatimusten valossa, mikä antaa arvokasta tietoa siitä, miten jatkossa tulisi kehittää vastaavanlaisia palveluita. Lisäksi yksittäisen palvelun käytöstä voidaan pitkän ajan kuluessa kerätä kokemuksia, joita voidaan jatkossa hyödyntää joko palvelukokonaisuuden muiden palveluiden tai muun toiminnan kehittämisessä. Valittu palvelu pyrittiin toteuttamaan siten, että siitä saadaan mahdollisimman paljon myös suoraa hyötyä kaupunkisuunnittelun ja osallistamisen tarpeisiin.

### 4.1 *Palvelukokonaisuuden kuvaus*

Palvelukokonaisuus koostuu yksittäisistä palveluista, jotka analysoivat paikkatietojärjestelmässä olevaa tekstimuotoista kommenttiaineistoa. Palvelut tukevat toisiaan siten, että ne tarjoavat erityyppisiä kuvakulmia aineiston sisältöön. Kaikki palvelut toimivat samalla alustalla, Kehittämisfoorumin paikkatietojärjestelmässä, ja käyttäjä voi siirtyä käyttämään toista palvelua hyödyntäen edellisestä saamiaan tuloksia. Reagoimalla kustakin analyysikerrasta saamiinsa tuloksiin ja käyttämällä eri palveluita iteratiivisesti, käyttäjälle muodostuu selkeä ja oikeellinen kokonaiskuva aineiston sisällöstä.

Palvelujen tarkoituksena ei ole antaa valmiita vastauksia kaupunkisuunnittelijan kysymyksiin, vaan käyttäjän on edelleen luettava ainakin joitakin kommentteja ymmärtääkseen niiden merkityksen. Järjestelmä auttaa käyttäjää hahmottamaan minkä asioiden ympärillä keskustelu liikkuu ja missä paikoissa ja kuinka paljon eri aiheista



puhutaan.

Palvelukokonaisuuden ytimenä toimii kaikille palveluille yhteinen tekstianalyysiohjelma, joka muodostaa aineistosta aiheavaruusmallin. Abstraktissa moniulotteisessa aiheavaruudessa kullakin kommentilla on sen sanastoon perustuva sijainti. Aiheavaruuden kukin ulottuvuus vastaa yhtä puheenaihekokonaisuutta. Aiheavaruus voidaan laskea erikseen riippumatta palveluiden käytöstä ja palvelut voivat nopeasti hyödyntää esikäsiteltyä aineistoa omissa analyyseissään, jotka puolestaan käynnistetään vain käyttäjän toimesta.

Kappaleessa 3.3 listatut vaatimukset soveltuvat kaikille palvelulle, tosin tiettyjen palveluiden kohdalla voi olla vaikeampi saavuttaa tiettyjä kohtia.

#### **4.1.1 Puheenaiheet tietyssä kartan pisteessä**

Puheenaiheet tietyssä kartan pisteessä esittävä palvelu yhdistää paikka- ja puheenaihetiedon siten, että käyttäjän valitsemalle maastonkohdalle etsitään kyseistä kohtaa parhaiten kuvaavat aihe sanat.

Käyttäjä valitsee paikkatietojärjestelmän kartalta hiirellä maastonkohdan tai alueen, ja palvelu palauttaa käyttäjälle tiedon siitä, mistä puheenaiheista kyseisessä paikassa tai alueella puhutaan. Täsmälleen kyseisessä kohdassa tai alueella ei tarvitse olla ainuttakaan kommenttia, vaan palvelu antaa tuloksen sen mukaan, mitkä aiheet ympäröivissä tekstinäytteissä nousevat voimakkaimmin esiin.

Karttapisteen sijainti aiheavaruudessa voidaan approksimoida esimerkiksi interpoloimalla ympäröivien pisteiden sijainteja aiheavaruudessa. Tuloksena saatava puheenaihe voidaan esittää kyseistä aiheavaruuden kohtaa parhaiten kuvaavien yksittäisten sanojen perusteella käyttämällä sana-avaruudelle tehtävän pääkomponenttianalyysin sivutuotteena saatavaa muunnosmatriisia, josta löytyy kunkin sanan painokertoimet jokaisen pääkomponentin suhteen.

Palvelu voi valita näytettävät puheenaiheet siten, että se antaa kaikille kommentteille

painokertoimen, joka voi olla esimerkiksi kääntäen verrannollinen etäisyyteen valitusta maastonkohdasta. Sen jälkeen se laskee kaikkien näytteiden aihevektorien painotetun keskiarvon. Tuloksena saadaan approksimaatio valitun maastonkohdan puheenaiheesta. Aiheavaruudesta voidaan sen jälkeen etsiä kyseistä aihetta lähimmät sanat esimerkiksi euklidisen etäisyyden tai pienimmän avaruuskulman mielessä. Palvelu näyttää tuloksena vakiomäärän parhaita aihe sanoja tai kaikki sanat, jotka ovat tietyn etäisyyden päässä saadusta aiheavaruuden pisteestä.

Palvelun avulla käyttäjä voi selvittää mikä asia puhuttaa kommentoijia eniten tietyssä paikassa. Kaupunkisuunnittelija voi hyödyntää tietoa esimerkiksi selvittäessään tietyn paikan ominaispiirteitä tai arvioidessaan asukkaiden suhtautumista niihin.

#### **4.1.2 Samankaltaisten kommenttien tunnistus**

Keskenään samankaltaiset kommentit voidaan löytää helposti niiden aihevektoreita vertailemalla. Ominaisuutta voidaan hyödyntää siten, että käyttäjän lukiessa yhden kommentin sisältöä hänelle esitetään lista muista kommenteista aiheen samankaltaisuuden mukaan paremmuusjärjestyksessä. Tällöin käyttäjä voi helposti siirtyä lukemaan muita samankaltaisia kommentteja ja saada selville samaan aiheeseen liittyviä asioita, jotka eivät selviä yhtä kommenttia lukemalla. Palvelun avulla kaupunkisuunnittelija saa paremman kuvan siitä, onko hänen lukemansa kommentti yksittäinen mielenilmaus, vai koskettaako asia paikallisia toimijoita yleisemminkin.

Palvelun toimintaa voidaan tehostaa näyttämällä aihepiiriltään lähimmät kommentit korostettuina kartalla. Tällöin käyttäjä saa selville myös puheenaiheen spatiaalisen luonteen, eli sen, missä paikoissa ja kuinka laajalla alueella aiheesta puhutaan.

#### **4.1.3 Klusterointi**

*Klusteroinnilla* tarkoitetaan informaatiotekniikassa näytteiden automaattista ryhmittelyä niiden piirteiden samankaltaisuuden mukaan (Inaba *et al.* 1994). Kommenttiaineiston klusterointi niiden puheenaiheiden mukaan tarjoaa käyttäjälle helpon sisäänkäsyn aineiston tutkimiseen. Käyttäkseen palvelua käyttäjän ei välttämättä tarvitse tietää

mistä aiheesta aineisto kertoo, olla lukenut ainuttakaan sen kommenttia, tai tietää edes yhtäkään sanaa jota aineistosta löytyy. Eri ryhmiin kuuluvat kommentit voidaan esittää kartalla värillisinä ryhminä. Lisäksi kustakin ryhmästä voidaan tuoda korostetusti esille yksittäinen kommentti, joka kuvaa sisällöltään parhaiten koko klusterin aihepiiriä. Tällainen edustusnäyte voidaan etsiä klusterin keskipisteen läheltä aiheavaruudessa. Kommentin sijaan klusterin sisältöä voidaan valita edustamaan myös yksittäisiä aiheavaruuden osaa kuvaavia aihe sanoja. Klustereiden tunnistuksen helpottamiseksi aihe sanat voidaan esittää kartalla klusterin rajaaman alueen sisällä.

Klusterointimenetelmiä on moninaisia. Ne pyrkivät löytämään yhtäläisyyksiä ja eroavaisuuksia aineistosta siten, että klustereiden väliset rajat ovat mahdollisimman mielekkäitä. Usein tämä ei kuitenkaan ole yksiselitteistä, vaan toimintaan vaikuttaa valittu klusterointikriteeri. Valinta voidaan kuitenkin tehdä etukäteen siten, että uuden käyttäjän ei tarvitse sitä osata tehdä. (Inaba *et al.* 1994)

Aineiston klusterirakennetta tarkastelemalla käyttäjä saa kattavan yleiskuvan koko aineiston eri puheenaiheista, minkä jälkeen hän voi ryhtyä tarkastelemaan häntä kiinnostavia aiheita yksityiskohtaisemmin muiden analyysipalveluiden avulla.

#### **4.1.4 Alueen rajaaminen puheenaiheen mukaan**

Kartalla olevalle pisteelle voidaan löytää sitä vastaava puheenaihe aiheavaruudessa. Vastaavasti aiheavaruuden yksittäiselle pisteelle voidaan löytää sijainti kartalta. Sijainti ei kuitenkaan ole yksikäsitteinen, koska kyseinen aihe voi olla relevantti useassa kartan pisteessä.

Alueen puheenaiheen mukaan rajaavassa palvelussa käyttäjä antaa syötteenä puheenaiheen. Puheenaiheelle löytyy aiheavaruudesta yksikäsitteinen sijainti pääkomponenttianalyysin muunnosmatriisin avulla. Kommenttiaineistosta voidaan etsiä puheenaihetta lähimmät kommentit aiheavaruudesta esimerkiksi euklidisen etäisyyden tai avaruuskulman mukaan. Tulos voidaan pistejoukon sijaan esittää yhtenäisinä alueina jatkamalla kunkin aiheesta kertovan kommentin aluetta tietyllä etäisyydellä kommentista kappaleessa 2.3.2 kuvatulla tavalla ja yhdistämällä vierekkäiset alueet. Aiheeseen

kuulumattomat kommentit halutaan kuitenkin rajata alueen ulkopuolelle.

Palvelun avulla käyttäjä näkee kartalta missä paikoissa häntä kiinnostava puheenaihe on puhuttanut ihmisiä. Näin hän voi paikantaa esimerkiksi tietyn tyyppisiä tunteita herättäviä, tai tietyn ominaispiirteen omaavia paikkoja. Lisäksi käyttäjä voi tutkia palvelun avulla hänen itse jättämänsä kommentin aihepiirin sijoittumista kartalla.

Palvelun tuloksen tarkastelun helpottamiseksi käyttäjälle esitetään myös hakuaiheeseen kuuluvien kommenttien tekstisisällöt relevanssin mukaan paremmuusjärjestyksessä. Näin käyttäjä pääsee suoraan käsiksi kommenttien sisältöihin ja voi löytää parempia teemaan liittyviä hakuaiheita.

## ***4.2 Toteutettavaksi valitun palvelun menetelmä***

Diplomityön aikataulussa voitiin toteuttaa vain yksi yllä kuvatuista palveluista. Palvelu toteutettiin ”proof-of-concept” -hengessä, eli sen tarkoituksena on osoittaa, että tämän tyyppinen palvelu voidaan ylipäänsä toteuttaa, ja että asetetut tavoitteet voidaan saavuttaa. Toteutetusta palvelusta saatuja kokemuksia voidaan myöhemmin hyödyntää muita palveluja toteutettaessa. Työn tuloksella haluttiin myös samalla olevan mahdollisimman paljon arvoa itsessään. Sen vuoksi valittiin palvelu, josta arvioitiin saatavan eniten konkreettista hyötyä. Toteutettavaksi valittiin kappaleessa 4.1.4 kuvattu **alueen rajauksen puheenaiheen mukaan** toteuttava palvelu.

### **4.2.1 Tekstianalyysi**

Tekstianalyysin toteutustavaksi valittiin latentti semanttinen indeksointi, koska sen on todettu soveltuvan hyvin vapaamuotoisen tekstiaineiston luokitteluun. Vastaaviin tuloksiin olisi luultavasti päästy myös ainakin riippumattomien komponenttien analyysillä tai itseorganisoituvalla kartalla. Näistä vaihtoehdoista pääkomponenttianalyysin toteuttaminen arvioitiin kuitenkin teknisesti helpoimmaksi ja työn puitteissa toteuttamiskelpoisimmaksi.

Latenttiin semanttiseen indeksointiin liittyvä pääkomponenttianalyysi antaa tuloksena

näytteiden sijainnit aiheavaruudessa, sekä *muunnosmatriisiin*. Minkä tahansa tekstin sijainti aiheavaruudessa voidaan selvittää kertomalla tekstin sanavektori oikealta puolelta kyseisellä muunnosmatriisilla. Siten myös käyttäjän antama hakutermin voidaan viedä aiheavaruuteen, jolloin voidaan vertailla sen samankaltaisuutta kommenttien aiheisiin. Kuten kappaleessa 2.4.1 todettiin, mitä lähempänä tekstinäytteet ovat toisiaan aiheavaruudessa, sitä enemmän niiden puheenaiheet muistuttavat toisiaan. Sama pätee myös hakutermiin, vaikka se ei ollutkaan mukana aiheavaruuden virityksessä. Haun tuloksiksi valitaan hakutermiä lähimmät näytteet, koska ne käsittelevät parhaiten hakutermin kuvaamaa puheenaihetta.

Vektoriavaruusmallissa kommentin tai hakutermin sijainti aiheavaruudessa riippuu sen sanastosta. Kukin tekstin sisältämä sana voi vaikuttaa kommentin aiheeseen. Sana-avaruuden muodostuksessa voidaan käyttää monenlaisia painotusmenetelmiä, jotka eroavat toisistaan siinä, miten sanan esiintymien lukumäärä kyseisessä kommentissa, ja sanan esiintyminen koko aineistossa vaikuttavat matriisin arvoihin. Koska kirjallisuuden perusteella selkeää oikeaa ratkaisua ei löydy, kokeiltiin muutamia eri arvoja sopivan painotusmenetelmän löytämiseksi. Parhaisiin tuloksiin päästiin painottamalla sanat niiden frekvenssin mukaan kaavalla

$$w_i = \frac{1}{f_i + 5},$$

missä  $w_i$  on sanan paino ja  $f_i$  on niiden kommenttien määrä, joissa kyseinen sana esiintyy. Sillä, löytyykö sana näytteestä yhden vai useamman kerran, ei ole vaikutusta sanan painoon eikä näytteen sijaintiin sana-avaruudessa. Näytteen sanavektori muodostetaan asettamalla sen arvoksi kussakin siitä löytyvää sanaa vastaavassa ulottuvuudessa kyseisen sanan paino. Ulottuvuuksissa, joita vastaavaa sanaa ei löydy, sanavektori saa arvon 0. On siis vain kaksi vaihtoehtoa minkä arvon näyte voi saada kussakin ulottuvuudessa; joko  $w_i$  tai 0.

Dimension pienennyksessä säilytettävien pääkomponenttien määrä valittiin arvioimalla analyysin tuloksia eri arvoilla. Parhaisiin tuloksiin päästiin käyttämällä 86:tta pääkomponenttia.

## 4.2.2 Kommenttien siirto hyperpallon kehälle

Sana-avaruutta voidaan esikäsitellä ennen pääkomponenttianalyysia. Kappaleessa 3.3.3 asetetun vaatimuksen O1 mukaan aineistoa pitää käsitellä lähtökohtaisesti tasapuolisesti. Eri pituiset kommenttitekstit voivat saada sana-avaruudessa eripituisen vektorin, koska jokainen näytteestä löytyvä eri sana kasvattaa sanavektorin pituutta. Pitemmät vektorit vaikuttavat pääkomponenttianalyysin tulokseen enemmän kuin lyhyemmät, koska ne aiheuttavat varianssiin suuremman muutoksen. Lisäksi vektori yleensä lyhenee dimension pienennyksessä.

Aineiston tasapuolisen käsittelyn vuoksi kommenttien ja hakutermien sanavektorit skaalataan vakiomittaisiksi ennen pääkomponenttianalyysia. Tämä voidaan nähdä myös siirtona hyperpallon kehälle. Skaalaus tehdään jakamalla vektorin arvot pisteittäin vektorin pituudella. Näin jokaisen vektorin pituudeksi tulee 1.

Sama operaatio toistetaan dimension pienennyksen jälkeen, jotta kaikkien tekstien puheenaiheet saadaan yhtä etäälle origosta. Muunnos voidaan tulkita siten, että kunkin tekstin oletetaan kertovan omasta puheenaiheestaan yhtä voimakkaasti.

Hakutermit tehdään vastaava skaalaus ennen ja jälkeen pääkomponenttianalyysin, tosin operaatioiden lineaarisuudesta johtuen ensimmäinen normitus on täysin redundantti toimenpide. Hakutermien ja kommenttiaineiston yhtäläisen käsittelyn vuoksi se kuitenkin tehdään.

Kuvatulla menettelyllä saadaan kommentit, jotka jäisivät sanastonsa puolesta muuten heikkovoimaisiksi, esiin aineistosta ja pois aiheavaruuden keskiöstä. Keskiöön jäävät kommentit aiheuttaisivat ongelmia myös siinä tilanteessa, jossa hakutermejä ei pystyittäisi tunnistamaan valittujen pääkomponenttien avulla ja sekin jäisi aiheavaruuden keskelle. Tällöin hakutuloksena saataisiin lähinnä huonosti tunnistettuja kommentteja ja tulos olisi harhaanjohtava, koska teksteissä ei välttämättä puhuta samasta aiheesta.

### 4.2.3 Tekstiaineiston esikäsittely

Tekstianalyysi tehdään ainoastaan sanaston perusteella. Sanat tunnistetaan samaksi vain jos ne ovat täsmälleen samat, toisin sanoen jos niissä on samat kirjaimet samassa järjestyksessä. Vaikka latentti semanttinen indeksointi kykenee itsessään yhdistelemään sanojen merkityksiä, sen toimintaa voidaan parantaa huomattavasti esikäsittelemällä aineisto siten, että samaa tai lähes samaa tarkoittavia sanoja yhdistellään jo ennen sana-avaruuteen vientiä. Luonnollinen kieli, erityisesti suomen kieli, sisältää kuitenkin paljon taivutusmuotoja ja loppupäätteitä, minkä seurauksena samaa asiaa tarkoittavat sanat saavat eri muodon. Sanoja muokkaamalla saadaan pienennettyä sanojen lukumäärää ilman, että kadotetaan kovinkaan suurta määrää analyysin kannalta hyödyllistä informaatiota.

Kommentit esikäsitellään ennen sana-avaruuteen vientiä siten, että kaikki niiden sanat katkaistaan, ja käytetään vain sanojen vartaloita. Lisäksi isot kirjaimet muutetaan pieniksi kirjaimiksi ja aineistosta poistetaan välimerkit. Numeroita käsitellään kuitenkin kuten kirjaimia, myös sanojen osina.

### 4.2.4 Alueen muodostuksessa käytetty menetelmä

Puheenaiheen kattaman alueen muodostamiseksi käytetään tietoa näytteiden sijainneista sekä siitä, mitkä näytteistä kuuluvat positiivisiin hakutuloksiin. Eräs yksinkertainen ratkaisu olisi kuvata valitut näytteet laajempina ympyrän muotoisina alueina, jolloin lähekkäin olevista positiivisista hakutuloksista muodostuisi katkeamaton yhtenäinen alue. Tämä tapa ei kuitenkaan huomioisi mahdollisesti niiden väliin jääviä negatiivisia hakutuloksia, vaan tulos näyttäisi siltä, että alueen kaikissa palautteissa olisi puhuttu kyseisestä puheenaiheesta. Tämän välttämiseksi on alueen muodostuksessa huomioitava myös negatiiviset hakutulokset.

Haluttuun tulokseen päästään jakamalla koko ympäristö alueisiin sen mukaan, minkä näytteen vaikutuksen alle kukin maaston kohta voimakkaimmin kuuluu. Tuloksena esitetään vain hakutulokseen kuuluvien kommenttien alueet, jolloin siihen kuulumattomat kommentit rajautuvat alueen ulkopuolelle. Näin käyttäjä saa paremman

yleiskuvan siitä miten puheenaiheet jakautuvat alueella, koska vain samaan puheenaiheeseen kuuluvista vierekkäisistä näytteistä muodostuu katkeamaton alue. Alueiden muodostumisen perusteeksi valittiin kappaleessa 2.3.2 kuvattu Voronoi-diagrammi yhdistettynä alueen rajoittamiseen korkeintaan tietylle etäisyydelle kommentista.

Käytetyllä tekstianalyysimenetelmällä ei voida selvittää kuinka laajaa aluetta yksittäinen kommentti kommentoi. Ei ole myöskään mielekasta ajatella, että hyvin kaukana koko tarkasteltavasta alueesta keskusteltaisiin samasta aiheesta kuin palautteessa, joka on reunimmaisena siinä suunnassa. Kommenttien paikannuksen epätarkkuuden ja alueiden yhtenäisyyden vuoksi alueen leikkauksessa käytettävän ympyrän säteen on oltava riittävän suuri. Liian isolla säteellä alueista taas tulee omituisen muotoisia ja hyvin eri kokoisia. Mielekkäisiin tuloksiin päästiin käyttämällä alueen leikkaamiseen säteeltään noin sadan metrin kokoisia ympyröitä. Yli sadan metrin päässä lähimmästä kommentista oleva maastonkohta ei siis kuulu yhdenkään näytteen alueeseen, eikä siis voi koskaan tulla esitetyksi analyysin tuloksessa.

Menetelmässä yksittäinen alue jää sitä pienemmäksi, mitä lähempänä se on naapureitaan ja mitä enemmän se on niiden ympäröimä. Tämä olisi ristiriidassa tasapuolisuusvaatimuksen O1 kannalta, jos tasapuolisuudesta haluttaisiin pitää ehdottomasti kiinni. Alueen sädetä rajoittamalla saadaan myös kommenttien käsittelyä muokattua myös tasapuolisempaan suuntaan, kun yhdenkään kommentin kattama pinta-ala ei voi koskaan ylittää tiettyä raja-arvoa.

### ***4.3 Palvelun toiminnallisuus käyttäjän kannalta***

Palvelua käyttääkseen käyttäjän on avattava Kehittämisfoorumi-sivuston karttanäkymässä analyysipalvelun oma välilehti. Välilehdellä on tekstinsyöttökenttä, käynnistyspainike, tekstikenttä hakutuloksille sekä tyhjennyspainike. Karttakuva on myös koko ajan näkyvissä.

Palvelu käynnistetään käyttäjän toimesta. Käyttäjä kirjoittaa tekstinsyöttökenttään hakutermiä eli puheenaiheen, jonka kattaman alueen hän haluaa nähdä kartalla, ja painaa



käynnistyspainiketta. Kun analyysin tulos on valmis, kartalle ilmestyy näkyviin läpikuultava värillinen alue, joka vastaa haettua puheenaihetta. Lisäksi tekstikenttään ilmestyy hakutulosten otsikot sekä sisältötekstit numeroituina relevanssin mukaisessa järjestyksessä.

Alue esitetään kartalla omana kerroksenaan, joka näkyy kaiken kartalla olevan muun tiedon päällä. Kerroksen voi halutessaan kytkeä pois näkyvistä ja takaisin näkyviin kartan kulmassa olevasta kerrospaneelista. Lisäksi aiemmin haetut analyysitulokset voi siivota pois kartalta ja tekstikentästä tyhjennyspainiketta painamalla.

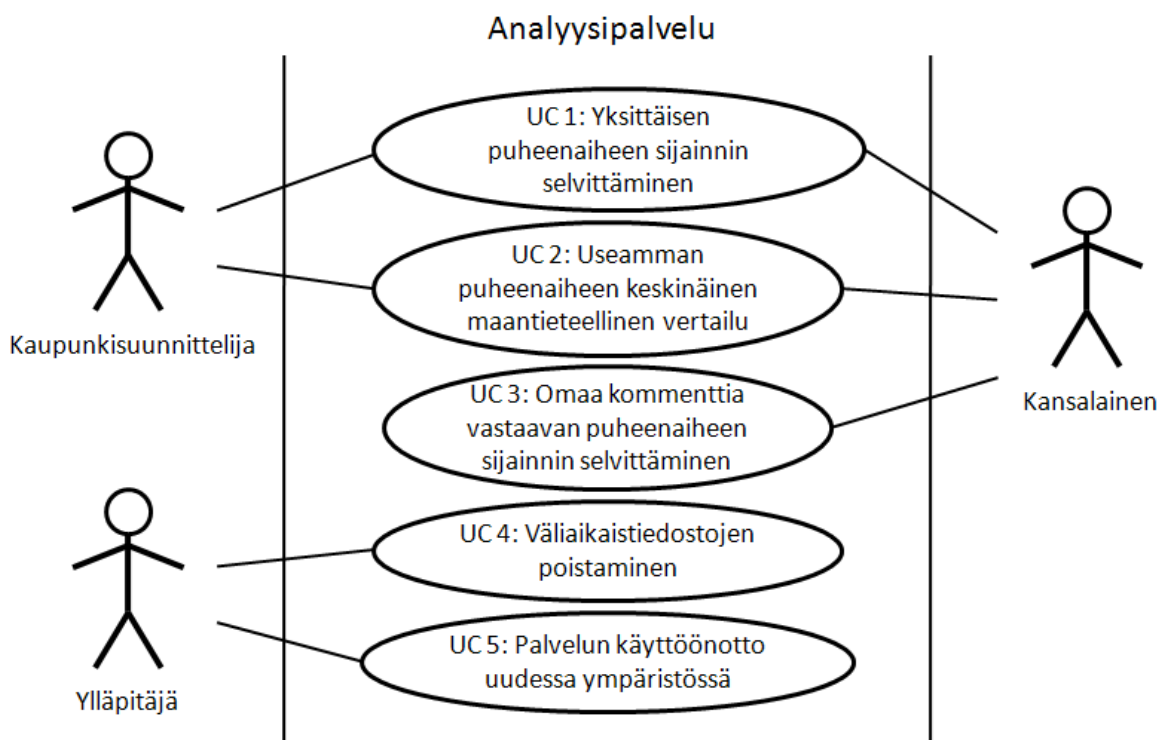
Useamman peräkkäisen haun tulokset näytetään kukin omalla värillään ja omassa kerroksessaan siten, että uusimman haun tulos näkyy aina päällimmäisenä. Käytössä on neljä eri väriä, joten neljän haun jälkeen aletaan käyttää järjestyksessä uudestaan samoja värejä.

Käyttäjä voi analyysin suorituksen aikana siirtyä muille välilehdille käyttämään paikallistietokartan muita palveluja, joiden avulla hän voi esimerkiksi asettaa kartalle näkyviin tai pois näkyvistä muuntyyppisiä paikkatietoja. Toiminta muilla välilehdillä ei vaikuta analyysipalveluun, vaan käyttäjä voi palata analyysivälilehdelle jatkamaan toimintaa siitä mihin se jäi välilehdeltä poistuttaessa. Analyysitulokset myös säilyy kartalla sellaisenaan muita välilehtiä käytettäessä mutta ei kuitenkaan karttanäkymästä poistuttaessa.

### **4.3.1 Käyttötapaukset**

*Käyttötapauksella* tarkoitetaan järjestelmälle ulkoisen elollisen tai elottoman toimijan suorittamaa tavoitteellista toimintaa, jonka seurauksesta on toimijalle hyötyä (Ambler 2004). Käyttötapaukset voidaan kuvata esimerkiksi *UML 2.0:n* käyttötapauskaaviolla. Tässä kappaleessa kuvataan toteutettavan palvelun käyttötapaukset.

Palvelun ulkoisia toimijoita ovat kaupunkisuunnittelijat, kansalaiset ja ylläpitäjät. Analyysipalvelun heille tarjoamat toiminnot on esitetty *UML 2.0 -käyttötapauskaavion* muodossa kuvassa 4. Toimintojen selitykset on listattu liitteen 1 taulukoissa 6 – 10.



**Kuva 4, Analyysipalvelun käyttötapauskaavio**

#### **4.4 Menetelmän asettamat vaatimukset**

Kappaleessa 3.3 listattujen vaatimusten todettiin asettavan sopivat rajaehdot valitun palvelun toteutukselle. Käytettävä tekstianalyysimenetelmä aiheuttaa kuitenkin erityisvaatimuksia kommenttiaineistolle. Tämän kappaleen lopussa olevaan taulukkoon 11 on listattu lisävaatimukset, jotka otetaan huomioon työn toteutuksessa ja arvioinnissa aiemmin mainittujen vaatimusten rinnalla.

Sanastoon perustuva tilastollinen tekstianalyysi vaatii yleensä aineistokseen suuren määrän tekstinäytteitä. Mitä enemmän näytteitä on, sitä enemmän näytteiden välisiä yhteyksiä löytyy, ja sitä pienempi vaikutus yksittäisillä satunnaisilla tekstinäytteillä on kokonaisuuden kannalta. Vektoriavaruusmallissa tämä on kriittistä aiheavaruuden

virittymisen kannalta.

Monet latentin semanttisen indeksoinnin edut voivat jäädä saavuttamatta, jos aineisto on liian vähäinen. Esimerkiksi sanojen merkitysten yhdenmukaistuminen, eli se, että sanaa käsitellään samankaltaisella tavalla kuin sen synonyymejä, ei välttämättä toteudu jos aineistosta ei löydy yhtäkään tekstinäytettä, jossa tietty sana esiintyisi synonyyminsä kanssa. Sanojen merkitykset voivat silti yhdistyä muiden sanojen vaikutuksesta, mutta vaikutus on sitä heikompi ja epävarmempi, mitä vähemmän näytteitä on.

Vastaavanlainen ongelma syntyy, jos tekstinäytteet ovat liian lyhyitä, eli jos niissä on liian vähän sanoja. Tällöin aineisto jää harvaksi ja todennäköisyys, että eri näytteistä löytyisi samoja sanoja jää pieneksi, millä on vastaavanlaiset vaikutukset kuin jos näytteitä on vähän.

Pienessä aineistossa, jossa sanojen väliset yhteydet ovat harvassa, ne yhteydet, jotka kuitenkin satutaan löytämään, saavat kohtuuttoman suuren painoarvon. Tämä käy selvästi ilmi tapauksessa, jossa aineiston joukkoon on päässyt sama tekstinäyte kahteen kertaan. Tällöin pääkomponenttianalyysi asettaa ensimmäisen pääkomponentin erittäin suurella todennäköisyydellä erottelemaan nämä duplikaattinäytteet muusta aineistosta. Duplikaattinäytteet saavat kyseisessä ulottuvuudessa korkean positiivisen tai negatiivisen arvon muun aineiston jäädessä vastakkaiselle puolelle lähelle origoa. Vasta seuraavat pääkomponentit erottelevat aineistoa halutulla tavalla, paitsi jos aineistossa on muitakin duplikaattinäytteitä.

Aineiston suuri määrä auttaa duplikaattiongelmaan siten, että ensimmäinenkin pääkomponentti voi asettua kuvaamaan aineiston muita piirteitä, jos aineiston muiden näytteiden väliltä löytyy riittävästi yhteyksiä. Duplikaatteja voidaan yrittää karsia aineiston esikäsittelyvaiheessa. Koska aineistoa ylläpidetään aktiivisesti, duplikaatit voidaan poistaa myös käsin, mikä on järjestelmän muunkin toiminnan kannalta paras ratkaisu.

Toisaalta liian suuri aineisto kasvattaa analyysin työtaakkaa. Sanamatriisin koko kasvaa huomattavasti näytteiden määrän kasvaessa, koska tällöin myös aineistosta löytyvien eri

sanojen määrä kasvaa. Tavoitteeksi voidaan ottaa, että järjestelmä toimii suurillakin aineistoilla, mutta sen on oltava riittävän nopea sellaisilla aineistoilla, joita käyttöympäristöihin oletetaan kertyvän.

Koska latentti semanttinen indeksointi on ohjaamaton menetelmä, aineiston aihepiirillä ei ole vaikutusta palvelun toimintaan. Palvelun pitäisi pystyä käsittelemään yhtä lailla mitä tahansa aihetta käsittelevää aineistoa. Siksi aineiston sisällölle ei aseteta lainkaan vaatimuksia.

*Taulukko 11, Vaatimukset käytettävän aineiston suhteen A1 – A3*

Vaatus A1	Sopivan kokoinen aineisto
Vaatus A2	Riittävän pitkät kommentit
Vaatus A3	Aineiston eheys

## 5 Tekninen ratkaisu

Tämän diplomityön puitteissa toteutettavaksi valittiin puheenaiheiden kattamat alueet visualisoiva tekstianalyysipalvelu. Toteutuksen alustana toimiva järjestelmä, Kehittämisfoorumi, määritteli paljon sitä millä teknologioilla työ toteutettiin. Palvelun tekniikka toteutettiin kokonaisuudessaan ilmaisilla vapaan lähdekoodin työkaluilla.

Palvelu toimii valmiin paikkatietojärjestelmän osana, joten se käyttää paljon järjestelmässä jo ennestään olevia resursseja, kuten karttanäkymää, tietokantaa aineistoinen, ja järjestelmää ylläpitävää palvelinta. Myös käyttöliittymä on jo ennestään olemassa, tosin siihen täytyy tehdä palvelua varten lisäyksiä. Järjestelmästä ei tarvitse poistaa mitään ja sisällönhallintajärjestelmän modulaarisuuden vuoksi siihen tehtävät muutoksetkin voidaan tehdä täysin sivuston normaalin ylläpitorajapinnan kautta koskematta sen lähdekoodiin.

Palvelun toteutusratkaisu on pohjimmiltaan kaksitasoinen asiakas-palvelin - arkkitehtuuri, jolla suoritetaan varsin perinteinen tietokantahaku. Tietokannasta haettavalle tiedolle tehdään kuitenkin tavanomaista raskaampi käsittely palvelimella analysointivaiheessa ennen kuin sen tulokset välitetään asiakaspääteelle.

Palvelun tekninen toiminta jakautuu karkeasti Internet-selaimessa toteutetun käyttöliittymän, palvelimella toimivan Drupal-modulin ja spatiaalisen tekstianalyysin suorittavan Perl-ohjelman toimintaan sekä niiden väliseen kommunikaatioon. Lisäksi kokonaisuuteen kuuluu oleellisena osana yhteys järjestelmän tietokantaan.

### 5.1 Palvelun toiminnan logiikka

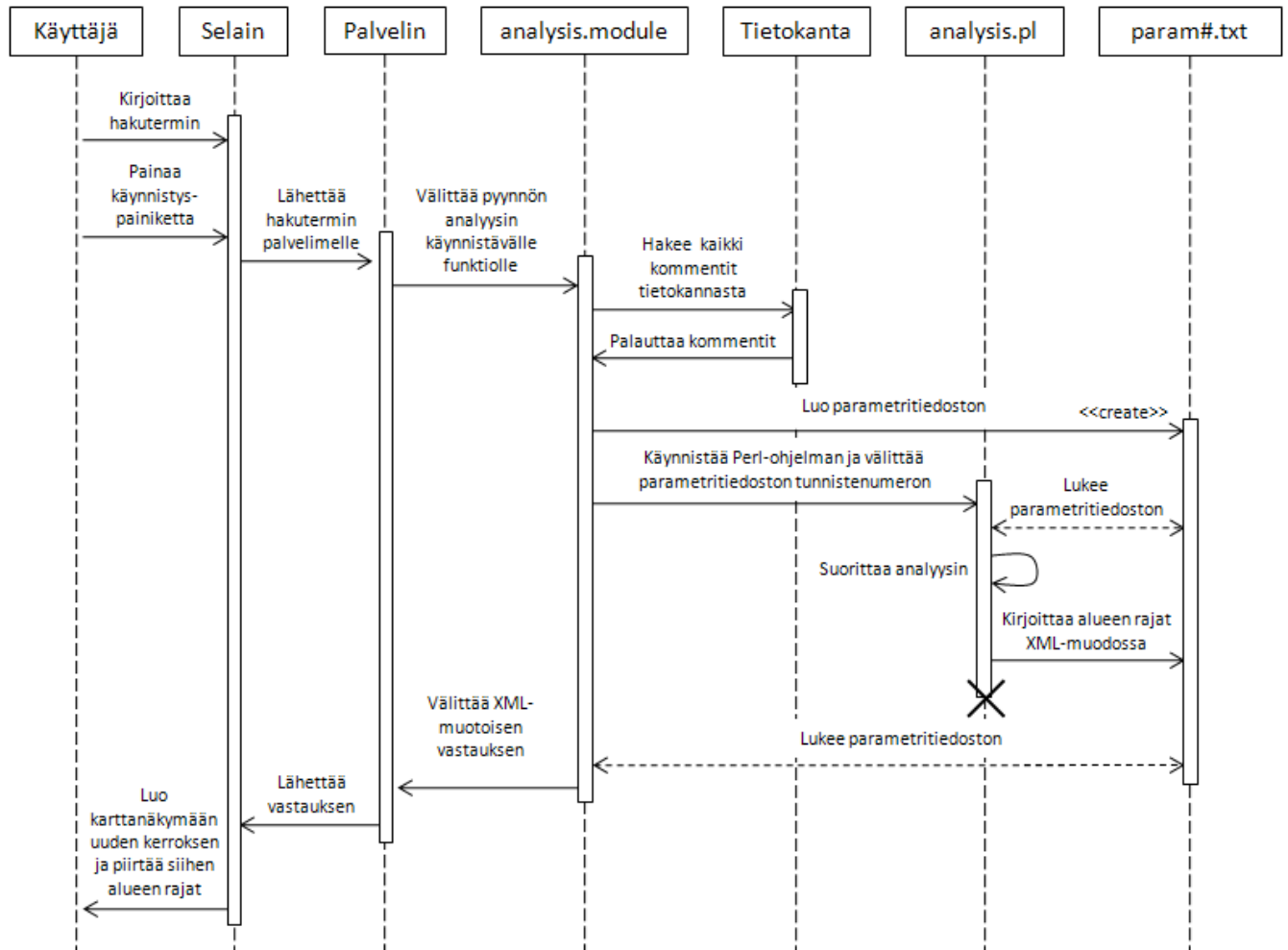
Käyttäjän kirjoittama hakutermi välitetään *XMLHTTP*-pyyntönä palvelimelle. Palvelin välittää pyynnön Drupal-sisällönhallintajärjestelmälle, joka puolestaan välittää hakutermiä parametrina *Analysis-modulissa* erikseen määritellylle funktiolle. Funktio hakee tietokannasta hakutermistä riippumatta koko kommenttiaineiston ja määrittää kyseiselle analyysi-instanssille oman satunnaisen tunnistenumeron. Funktio luo tekstitiedoston, jonka nimessä on kyseinen tunnistenumero, ja kirjoittaa tiedostoon

hakutermien sekä kommenttiaineiston koordinaatit ja tekstit.

Tämän jälkeen funktio käynnistää *Perl*-ohjelman ja välittää sille komentoriviparametrina analyysi-instanssin tunnistenumeron. *Perl*-ohjelma lukee käsiteltävän aineiston numeroa vastaavasta tekstitiedostosta ja suorittaa sen perusteella spatiaalisen tekstianalyysin. *Perl* kirjoittaa analyysin tuloksen XML-muodossa samaan tiedostoon ja lopettaa itsensä.

Analysis-modulin funktio saa lopetusilmoituksen *Perl*-ohjelmalta ja lukee tiedostosta analyysin tuloksen. Funktio pyytää palvelinta lähettämään analyysin tuloksen selaimelle. Tulos esitetään käyttäjälle järjestelmän karttanäkymässä geometrisessa muodossa omana kerroksenaan.

Palvelun tekninen toiminta on esitetty UML2.0 -sekvenssikaavion muodossa kuvassa 5. Sekvenssikaavio soveltuu erityisesti kuvaamaan monimutkaisen kokonaisuuden logiikkaa esittämällä sen osien välisen kommunikaation aika-akselilla yhden käyttötapauksen osalta (Ambler 2007). Tähän kaavioon on valittu kappaleessa 4.3.1 esitetyistä palvelun käyttötapauksista ensimmäinen, UC1, Yksittäisen puheenaiheen sijainnin selvittäminen, joka kuvaa järjestelmän perustoimintaa ja laajimmalle käyttäjäkunnalle suositeltavaa käyttötapaa.



Kuva 5, Sekvenssikaavio analyysipalvelun toiminnasta

## 5.2 Järjestelmän sisäinen kommunikaatio

Selaimen, palvelimen ja tietokannan välinen kommunikaatio on toteutettu *AJAX* (Asynchronous JavaScript and XML) -menetelmällä. AJAX ei ole ohjelmointikieli, vaan eri ohjelmointikielten ja ratkaisujen kokonaisuus, joka mahdollistaa interaktiivisen sisällön luomisen Internet-sivustolle ilman, että sivua tarvitsee kokonaisuudessaan päivittää jokaisen muutoksen yhteydessä. Näin karttanäkymään voidaan esimerkiksi lisätä uusi analyysin tuloksen näyttävä kerros säilyttäen samalla muiden käyttäjän suorittamien toimenpiteiden, kuten karttanäkymän vierittämisen, tulokset.

Yhteys Drupal-modulista tietokantaan on toteutettu MySQL:n ohjelmallisella rajapinnalla Drupalin tarjoamien funktioiden kautta. Tietokantayhteys oli olemassa jo ennestään koska sitä käytetään paikkatiedon säilyttämiseen ja käsittelyyn, joten sen käyttöönotto uudessa modulissa oli helposti toteutettavissa. Palvelun toteutus ei vaatinut muutoksia tietokannan osalta, koska palvelu ainoastaan lukee kannassa olevaa tietoa. Palvelu ei tallenna tietokantaan mitään tietoa.

Perl-ohjelman käynnistäminen PHP:lla kirjoitetusta Drupal-modulista käsin sujuu helpoiten *exec*-funktiolla, joka sallii yksirivisen komentoriviparametrin välittämisen käynnistettävälle ohjelmalle. Koska koko aineistoa ei voida näin välittää, se kirjoitetaan väliaikaisesti tiedostoon ja parametrina välitetään vain tunnistenumero, jolla Perl-ohjelma löytää oikean tiedoston. Samoin Perlistä käsin ei voida palauttaa analyysin tulosta suoraan PHP:lle, vaan tulos on kirjoitettava tiedostoon. Yleensä on parempi käyttää input- ja output-tiedostoina eri tiedostoja, mutta Perl-ohjelmalle ei voitu teknisistä syistä antaa uuden tiedoston luomiseen tarvittavia oikeuksia.

### **5.3 Käyttöliittymämuutokset**

Drupalin modulaarisuus mahdollistaa pienten käyttöliittymämuutosten vaivattoman toteuttamisen, kunhan Drupalin oma toiminta on ohjelmoijalla ensin selvillä. Mikä tahansa yksittäinen moduli voi lisätä periaatteessa mihin tahansa sivuston näkymään toiminnallisuutta, joka automaattisesti kytkeytyy pois päältä ja näkyvistä kun modulin poistaa käytöstä.

Analyysimoduli lisää paikallistietokarttanäkymään oman välilehden, jossa on analyysin käyttöliittymän vaatimat painikkeet ja tekstikentät. Omalla välilehdellään niille ei tarvitse erikseen varata tilaa. Sivu venyy tarvittaessa alaspäin siten, että ne ovat aina näkyvissä ja käytettävissä. Pienellä resoluutiolla näkymää voi joutua vierittämään alaspäin, jotta ne saa näkyviin.

Tekstikentät ja painikkeet määritettiin Drupalin *lohko*- (engl. "block") rakenteella. Tekstikentät ovat html:n *form*-, ja painikkeet *button*-muotoisia. Käynnistuspainiketta painettaessa käynnistetään JavaScriptillä toteutettu funktio, joka välittää



tekstinsyöttökentässä sillä hetkellä olevan hakutermien XMLHTTP-pyyntönä palvelimelle. Myös tekstimuotoisten tulosten esittäminen toisessa tekstikentässä sekä vanhojen analyysitulosten poistaminen on toteutettu JavaScriptillä.

Tämän luvun lopussa on esitetty analyysipalvelun käyttöliittymänäkymä kuvassa 6.

## **5.4 Tekstianalyysi**

Palvelun tekstianalyysiosuus toteutettiin Perlillä, koska Perl on tehokas ohjelmointikieli tekstin käsittelyssä. Perliin on saatavilla laajennos, *Perl Data Language* (PDL), jonka avulla sillä voi helpommin tehdä monimutkaisia laskentaoperaatioita. PDL tukee matriisimuotoisia muuttujia ja osaa tehdä matriisikertolaskuja, joten se on sopiva apuväline latentin semanttisen indeksoinnin toteuttamisessa.

Yksittäistä palvelua toteutettaessa poikettiin suunnitelmasta tehdä täysin erillinen sanastoanalyysin tuottava ohjelma, joka toimisi palvelukokonaisuuden ytimenä, koska sen toteuttaminen olisi vaatinut liian paljon aikaa. Vastaaviin tuloksiin, ehkä vasteaikaa lukuun ottamatta, päästiin toteuttamalla sanastoanalyysi samassa yhteydessä kuin aineiston muukin käsittely. Latenttia semanttista indeksointia ei kuitenkaan kannata tehokkuussyistä ajaa joka analyysikerralla uudestaan, vaan noin kymmenen kertaa parempaan suorituskyykyyn päästään, kun analyysin tuloksena saatavan aiheavaruuden sisältö tallennetaan tekstitiedostoon, josta se voidaan myöhemmillä analyysikerroilla lukea.

Aiheavaruuden viritys pysyy samana analyysikerrasta toiseen, mikäli aineisto ei muutu. Palvelu ohjelmoitiin siten, että se havaitsee aineiston muuttumisen siinä olevien sanojen lukumäärän muutoksesta. Tämä ei ole täysin aukotonta, joten lisäehdoksi tekstitiedoston käytölle asetettiin, että tiedosto ei saa olla yli kuukauden vanha. Ylläpitäjä voi tarvittaessa myös käsin tuhota palvelimella olevan väliaikaistiedoston, jolloin laskenta tehdään seuraavalla ajokerralla uudestaan.

### 5.4.1 Sanan katkaisu

Sanan katkaisuun, eli sen muuntamiseen pelkäksi vartalokseen, löytyi useita vaihtoehtoisia teknologioita. *Voikko* ja *Finflect* ovat suomalaisia ohjelmia, joilla sanoja voi taivuttaa eri muotoihin, ja ne molemmat tunnistavat myös sanan vartalon. *Snowball* on Perliin asennettava lisämoduli, jota voi käyttää suoraan Perl-ohjelmakoodista käsin, ja sillä voi muuttaa sanat vartaloikseen useilla eri kielillä. Työssä valittiin käytettäväksi *Snowball*, koska se ei vaadi muiden ohjelmointikielten käyttöä, eikä erillisten ohjelmien käynnistämistä. Lisäksi se on avoimen lähdekoodin ohjelma, joten sen käytöstä osana projektia ei voi seurata tekijänoikeuskiistoja.

Kommenttiaineiston sanojen katkaisu tehdään ennen sana-avaruuden muodostamista. Hakutermin sanojen katkaisu tehdään ennen sen vientiä sana-avaruuteen.

## 5.5 Alueiden muodostus

Analyysin tulos esitetään käyttäjän selaimessa karttanäkymän yhdessä omassa kerroksessaan. Tulos esitetään konveksien monikulmioiden joukkona siten, että vierekkäin olevista monikulmioista muodostuu katkeamattomia alueita, jotka näyttävät yhtenäisiltä ja siten yksittäisiltä monikulmioilta. Käyttäjän ei ole tarkoitus havaita vierekkäin olevien hakutulokseen kuuluvien alueiden välistä rajaviivaa, vaan mieltää nämä yhtenäiseksi alueeksi.

Alueiden muodostuksen perustana on Voronoi-diagrammi, joka on toteutettu *QHull*-nimisellä ohjelmistolla. *QHull* on *C*-kielinen ohjelma, jota voi käyttää sen omalla erillisellä käyttöliittymällä tai komentorivimuodossa. Perlin kautta *QHullia* käytetään komentorivimuodossa antaen sille samalla syötteenä käsiteltävä aineisto. Perl lukee *QHullin* antaman tuloksen erillisestä tekstitiedostosta.

*QHull* sisältää monia avaruudelliseen laskentaan erikoistuneita ohjelmia, joista kommenttien spatiaalisen analyysin toteuttamiseen käytettiin kahta, *QVoronoi* ja *QHalfia*.

QVoronoi muodostaa aineiston perusteella sen Voronoi-diagrammin. Ohjelmalle annetaan parametrina kaikkien kommenttien koordinaatit ja se palauttaa pisteitä vastaavat Voronoi-alueet soluja rajoittavien puoliavaruuksien yhtälöinä. Syötteenä annetaan myös hakutulokseen kuulumattomien kommenttien koordinaatit, koska niidenkin halutaan vaikuttavan alueen muotoon.

QHalf laskee konveksien monikulmioiden leikkauksen. Ohjelma käynnistetään kerran erikseen jokaista hakutulosta kohden. Sille annetaan parametrina kommentin Voronoi-alueita rajoittavat puoliavaruudet, sekä vakiosäteistä ympyrää approksimoiva puoliavaruuksien joukko. QHalf palauttaa tuloksena monikulmion kulmapisteet, jotka kuvaavat syötettyjen puoliavaruuksien, ja siten myös Voronoi-alueen ja ympyrän, leikkauksen. Leikkauksena saatavaa konveksia monikulmiota käytetään kommentin kattamana alueena.

QHalf palauttaa kulmapisteet epämääräisessä järjestyksessä, joten ne täytyy järjestää ennen kuin niistä voi muodostaa oikeanlaisen monikulmion. Sekä Voronoi-alue että ympyrä ovat konvekseja kuvioita. Konveksin monikulmion määritelmästä (Anon. 2010) seuraa, että kahden konveksin kuvion leikkaus on aina myös konvekksi. Konveksin kuvion kulmapisteiden astekulmat ovat keskipisteestä katsottuna vastapäivään kuljettaessa nousevassa järjestyksessä, joten kulmapisteet voidaan järjestää taulukkoon tämän perusteella yksiselitteisesti.

Alueiden koordinaatit kirjoitetaan XML-muotoon siten, että yhden elementin sisällä on aina yhden kommentin kattaman alueen määrittävän monikulmion kulmapisteiden koordinaatit. XML-muotoinen esitys kirjoitetaan syötetiedostoon Drupal-modulin luettavaksi, ja Perl-ohjelma päättyy.

## **5.6 Visualisointi**

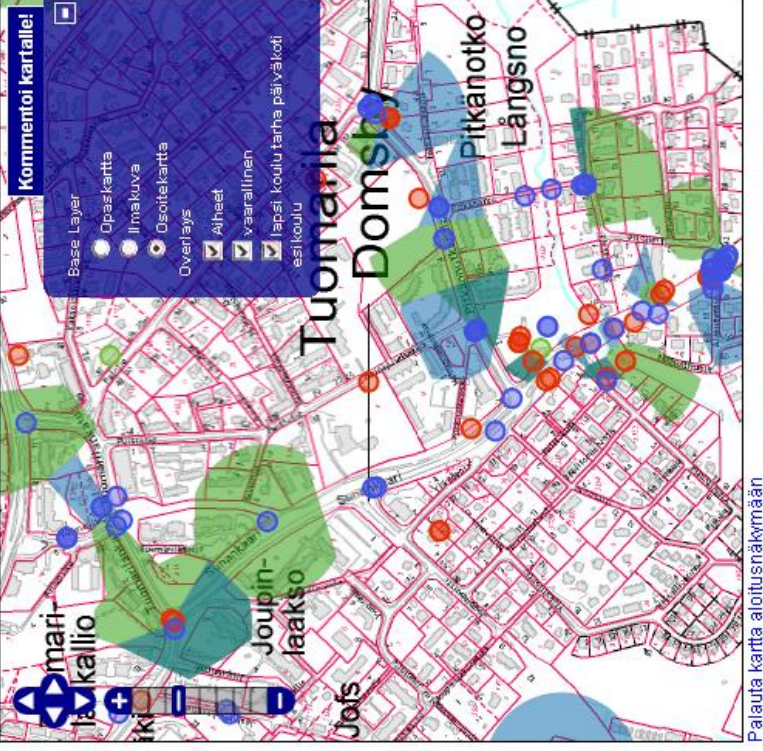
Analyysin tuloksen esittäminen paikkatietojärjestelmän karttanäkymässä on toteutettu JavaScriptillä. XMLHTTP-kutsua lähetettäessä määritetään funktio, joka käynnistyy automaattisesti vastaussanomien saapuessa. AJAX:n mukaisesti visualisointi tapahtuu asynkronisesti, eli käyttäjän toiminnalla palvelun käynnistymisen jälkeen ei ole vaikutusta

tuloksen esittämisen kannalta, olettaen että käyttäjä ei kuitenkaan siirry selaimellaan pois paikallistietokartan näyttävältä sivulta.

Visualisointi tapahtuu lisäämällä kartalle päällimmäiseksi uusi kerros, johon piirretään XML-muodossa saadut monikulmiot OpenLayersin funktioita käyttäen. Monikulmio muodostetaan *LinearRing*-muodossa käyttäen sen kulmapisteinä *Point*-muotoisia koordinaattitietoja.

Jotta puheenaiheita kuvaavat alueet voi erottaa toisistaan, niiden ulkoasun on oltava erilainen. Eräs havainnollinen tapa tehdä alueista eri näköisiä on värittää ne eri väreillä (Ingram & Benford 1995). Kutakin yksittäistä analyysikertaa vastaa oma värinsä, jolla alueet piirretään. Alueet piirretään hieman läpikuultavina siten, että niiden opasiteetti, ja vastaavasti myös läpikuultavuus, on 50 %. Koska kyseinen kerros näkyy kartalla päällimmäisenä, läpinäkyvyys mahdollistaa sen, että käyttäjä näkee hyvin myös alueen alla olevat kohteet, kuten itse kommenttien sijaintipisteet. Lisäksi, kuten kappaleessa 2.3.1 todettiin, läpikuultavuus soveltuu hyvin epämääräisen informaation visualisointiin.

Kerros jää karttanäkymään näkyviin, mutta käyttäjä voi piilottaa sen pois näkyvistä OpenLayersin automaattisesti tarjoaman kerrospaneelin kautta. Kukin kerros on nimetty sitä vastaavan hakutermien mukaisesti. Käynnistämällä analyysin uudestaan eri hakuterminä, käyttäjä saa kartalle näkyviin uuden kerroksen, jonka tulosta voi vertailla aiempien hakujen tuloksiin. Läpikuultavuuden ansiosta eri hakukertojen värit näyttävät päällekkäin osuessaan kyseisten värien sekoitukselta, minkä on tarkoitus saada käyttäjä intuitiivisesti mieltämään alue molempiin hakutermeihin liittyväksi.



Palauta kartta aloitusnäkyään

Valitse aihe Hae kommentteja Näytä listana

Analysoi

Kirjoita aiheasanat:  
[lapsi koulu tarha päivät Analysoi]

1. Vaarallinen suojatie  
Nopeusrajoitus 30 mutta kaahataan kovempaa, voisiko olla korotettu suojatie ??? Suojatie on esikoulun ja päiväkodin kulmalla ja koulualaisten reitti.

2. Tien ylitys  
Tällä kohtaa mm. 9/2007 pyöräilevä lapsi tuli auton töytäisemäksi suojatien kohdalla. Kovat nopeudet, Kauniaisten suunnasta mäennpyölyä estää näkyvyyden.

3. ei tien jatkoa

Poista vanhat analyysitulokset

- Etusivu
- Paikallistietokartta**
- Keskustelut
- Ajankolitaista
- Tapahtumat
- Uutiset
- Tiedotteet
- Ilmoituspalsta
- Ilmoita uutinen tai tapahtuma
- Kelittämiskohteet
- Kelittämisprojektin kuulumisia
- Virastokeskus
- Radan eteläpuoli
- Kirkkojärvi
- Espoonjoki
- Nimismiehenpelto
- Suvela
- Aluekirjasto
- Entresse
- Kaupunkipolut
- Asiaa Keski-Espoosta
- Lukijoiden blogi
- Kotitietunustoja
- Palvelut
- Alteen toimijat
- Kuvagalleria
- Ohje
- Palautte
- Yhteystiedot

Kuva 6, Analyysipalvelun käytöllittymä Kehittämisfoorumin paikallistietokarttanäkymässä

## 6 Työn arviointi

Työn kokeellisessa osuudessa toteutettua analyysipalvelua testattiin evaluoimalla käytettyä menetelmää sekä kokeilemalla palvelun toimintaa vapaamuotoisemmin. Koekäyttäjänä toimi Vantaan kaupungilla kaupunkisuunnittelijana työskentelevä arkkitehtuurin ja käytettävyyden asiantuntija. Palvelua arvioidaan vertaamalla testien ja koekäytön tuloksia kappaleissa 3.3 ja 4.4 esitettyihin ennalta asetettuihin vaatimuksiin.

### 6.1 Menetelmän evaluointi

Tekstianalyysin ja spatiaalisen analyysin toimivuuden toteamiseksi suoritettiin kaksi koetta, joissa tarkasteltiin lähemmin kahden eri hakutermien antamia tuloksia. Ensimmäisessä kokeessa hakuterminä käytettiin yksinkertaista ja konkreettista käsitettä ”pyörä”. Toisessa kokeessa hakutermiksi valittiin paikkaan kohdistuvaa aversiota eli vastenmielisyyttä kuvaavia sanoja. Sanat valittiin intuitiivisesti siten, että niitä voisi kuvitella käytettävän, kun halutaan kommentoida alueen luotaantyöntävyyttä. Hakutermiksi valittiin ”*likainen ankea ahdistava ruma hyi yäk allottävä epäsiisti*”.

Molempien kokeiden perusteella arvioidaan ensin käytetyn tekstianalyysimenetelmän kykyä löytää tiettyyn aihepiiriin kuuluvia kommentteja ja tekstiaineiston piileviä semanttisia rakenteita. Sen jälkeen arvioidaan palvelun maantieteellisten ominaisuuksien toimintaa, eli kykyä paikantaa tietyn tyyppisiä piirteitä omaavia alueita kartalta.

#### 6.1.1 Tekstianalyysin evaluointi

Ensimmäinen koe oli tarkoitus suorittaa jakamalla aineisto ensin käsin aihepiiriin liittyviin ja siihen liittymättömiin kommentteihin, ja sen jälkeen kokeilemalla antaako analyysipalvelu kyseisellä hakutermillä tuloksena aiheeseen liittyvät kommentit. Ongelmaksi koitui kuitenkin rajanveto aihepiiriin kuulumisen ja kuulumattomuuden välillä. Jos rajan vetää eri kohtaan kuin mihin analyysipalvelu sen vetää, tulokset ovat väistämättä huonoja. Tämän vuoksi tulokset jaettiin useampaan luokkaan sen mukaan kuinka laajassa mielessä ne liittyvät haettuun aiheeseen.

Hakuterminä käytettiin sanaa ”pyörä”. Kaikki polkupyöristä, pyöräilystä, pyöräilijöistä ja pyöräliikenteestä kertovat kommentit luokiteltiin aiheeseen eniten liittyviksi ensimmäisen luokan kommentteiksi. Kevyestä liikenteestä yleisesti kertovat kommentit luokiteltiin toisen luokan kommentteiksi, koska ne liittyvät pyöräilyn lisäksi yhtä lailla esimerkiksi jalankulkuun. Monissa hakutuloksissa puhuttiin lasten liikenneturvallisuudesta, joka on kevyeen liikenteeseen ja myös pyöräilyyn liittyvä asia, joten ne luokiteltiin kolmannen luokan kommentteiksi. Neljänteen luokkaan valittiin kaikki muut liikenteestä, kuten esimerkiksi autoliikenteestä, kertovat kommentit. Kommentit, jotka eivät liittyneet liikenteeseen, luokiteltiin viidenteen eli vähiten aiheeseen liittyvään luokkaan.

Luokittelun tarkoituksena ei ollut rinnastaa kaikkea liikenneaiheista keskustelua pyöriin liittyväksi, vaan antaa keinoja kokeen tuloksen tarkempaan arviointiin. Jako eri ryhmiin tehtiin pitkälti kommentteista löytyvän sanaston mukaan, ja monin paikoin jouduttiin tekemään valintoja sen suhteen, miten laajassa merkityksessä asiat nähdään, esimerkiksi kevyeen liikenteeseen liittyen. Monet jalankulun kysymykset, kuten suojateiden sijainnit, liittyvät yhtä lailla myös pyöräilyyn.

Analyysipalvelu antoi tuloksena 26 kommenttia. Näistä 8 oli ensimmäiseen luokkaan kuuluvia pyöräilystä kertovia kommentteja. 10 kertoi kevyestä liikenteestä, 11 lasten liikenneturvallisuudesta, ja 5 oli muita liikenteeseen liittyviä kommentteja. Vain yksi kommentti ei liittynyt lainkaan liikenteeseen. Tulokseen sisällyttämistä 182:sta kommentista 5:ssä mainittiin pyöräilyyn, tosin niistä neljässä puhuttiin pyöräilyn rinnalla myös jalankulusta.

Tavallisten tilastollisten tekstianalyysisovellusten yleiseen suorituskyykyyn verrattuna tulos ei näytä erityisen hyvältä. Mihin tahansa kohtaan rajan aihepiiriin kuulumisen ja kuulumattomuuden välillä asettaakin, joko hylkäämisvirheen tai hyväksymisvirheen määrä on suuri, aiheuttaen joka tapauksessa suuren kokonaisvirheen määrän.

Vapaasti kirjoitetun tekstin luokittelu aihepiiriin mukaan on kuitenkin suhteellisen vaikea tehtävä (Booker *et al.* 1999). Tehtävän vaikeutta voisi havainnollistaa pyytämällä kahta ihmistä luokittelemaan aineiston aiheeseen kuuluvien ja kuulumattomien joukkoon. Mitä

luultavimmin hekin päätyisivät keskenään erilaisiin jakoperusteisiin, ja siten myös eri tuloksiin.

Aihepiiriin kuulumista tulisikin tarkastella ennemmin jatkumona kuin kaksijakoisena luokitteluna. Tältä kannalta katsottuna se, että vain yksi positiivinen hakutulos oli täysin aiheeseen liittymätön, voidaan nähdä erittäin hyvänä tuloksena. Analyysipalvelu onnistui valitsemaan koko aineistosta 26 kommenttia, joista vain yksi ei liittynyt lainkaan aiheeseen. Tämän perusteella voidaan väittää, että palvelu on ainakin jollakin tasolla tunnistanut aihepiirin ja pystyy lajittelemaan aineistoa sen mukaan.

Varsinaisen luokittelutuloksen lisäksi kokeesta ilmeni muita kiinnostavia seikkoja. Pyörä-sanana eri taivutusmuodot yhdistetään samaksi sanaulottuvuudeksi jo aineiston esikäsittelyvaiheessa, mutta sanan yhteys muihin aiheeseen liittyviin sanoihin syntyy vasta lateraalisen semanttisen indeksoinnin vaikutuksesta. Tuloksista päätellen palvelu onnistui hyvällä menestyksellä tunnistamaan pyöräilyn, pyöräilijän ja polkupyörän aiheeseen liittyviksi sanoiksi. Pyöräliikenne-sanana sisältävää yksittäistä kommenttia se ei kuitenkaan ottanut mukaan hakutulokseen.

Menetelmä onnistui yhdessä tapauksessa tunnistamaan kommentoijan kirjoitusvirheestä huolimatta ”pyörilijä”-sanana sisältäneen kommentin aiheeseen liittyväksi. Toinen merkittävä onnistuminen oli hyväksyä kommentti, jossa puhuttiin kevyen liikenteen sijaan ”autottomista ihmisistä”. Kummassakaan kommentissa ei ollut havaittavissa selkeitä sanallisia yhteyksiä muihin aiheeseen liittyviin kommentteihin. Tulos viittaa vahvasti tekstiaineiston latenttien semanttisten rakenteiden olemassaoloon sekä käytetyn menetelmän kykyyn tunnistaa ja hyödyntää niitä.

Toisen kokeen tuloksia oli helpompi käsitellä, koska ne oli helpompi jakaa aiheeseen liittyviin ja liittymättömiin. Palvelu ilmoitti, että kommenttiaineistosta ei löytynyt lainkaan sanoja ”ahdistava”, ”hyi”, ”yäk” ja ”ällöttävä”. Kyseiset sanat eivät siis vaikuttaneet analyysitulokseen lainkaan, vaan hakuterminä olisi voinut yhtä hyvin olla vain ”likainen ankea ruma epäsiisti”. Vaikka puolet sanoista jäi hyödyntämättä, hakutermin aihealue säilyi suurin piirtein samana, ja analyysin tuloksen voitiin odottaa olevan hyvin samansuuntainen kuin jos kaikkia sanoja olisi voitu hyödyntää.



Tekstianalyysin tuloksena saatiin 14 kommenttia, joista 10:n katsottiin liittyvän aiheeseen ja 4:n ei. Hakutuloksen ulkopuolelle jäi 5 aiheeseen liittyvää kommenttia ja 191 aiheeseen liittymätöntä.

Tekstianalyysin havaittiin toimivan myös muiden kuin hakutermiin kirjoitettujen sanojen perusteella. Aiheeseen kuuluviksi valituissa kommentteista löytyi hakusanojen sijaan esimerkiksi seuraavia aiheeseen liittyviä termejä: ”roska”, ”roskis”, ”hälyttävä epäkohta”, ”huonokuntoinen”, ”surkea” ja ”kamala”. Latentin semanttisen indeksoinnin toiminnan kannalta on positiivista, että tunnistaminen voitiin tehdä myös ilman, että kommentista löytyi jokin käytetyistä hakutermeistä.

Tekstianalyysin tulosta voidaan arvioida käsitteillä *saanti* (engl. ”recall”) ja *tarkkuus* (engl. ”precision”). Saannilla tarkoitetaan aiheeseen kuuluvien kommenttien sisällymistä hakutulokseen ja tarkkuudella vastaavasti hakutulokseen sisältyvien kommenttien kuulumista aiheeseen. Tunnusluvut voidaan ilmaista aiheeseen kuuluvien hakutulosten prosentuaalisena osuutena kaikista aiheeseen kuuluvista kommentteista (saannin tapauksessa) tai kaikista hakutuloksista (tarkkuuden tapauksessa). Tuloksesta on tärkeää mitata molemmat tunnusluvut, koska tuloksen saa helposti manipuloitua näyttämään täydelliseltä, jos käytetään vain toista mittaria. (Wendy & Sundheim 1991)

Toisen kokeen saanniksi saadaan

$$\frac{10}{15} = 66,7 \%,$$

ja tarkkuudeksi

$$\frac{10}{14} = 71,4 \%.$$

Lukemille ei löydy kirjallisuudesta kunnollista vertailukohtaa, jossa olisi käytetty samantyyppistä aineistoa. Eräässä tekstianalyysimenetelmien kartoituksessa (Wendy & Sundheim 1991) vuodelta 1991 listattiin 18:n eri menetelmän saantia ja tarkkuutta. Näistä vain kolmella saanti oli yli 66,7 %, ja neljällä tarkkuus yli 71,4 %. Kartoituksessa

käytetty analyysitehtävä oli kuitenkin hieman erityyppinen ja kartoitus on tehty lähes 20 vuotta sitten, joten lukuja ei voida suoraan vertailla.

Tulokseen vaikuttaa myös se, kuinka suuri osa aineistosta on aiheeseen kuuluvia. Suoritetun kokeen tulos voidaan mieltää hyväksi, koska alun perin vain 7,2 % koko aineistosta oli aiheeseen kuuluvia, ja niistä suurin osa tunnistettiin oikein siten, että hakutulokseen päätyi mukaan vain muutama aiheeseen kuulumaton kommentti.

### **6.1.2 Spatiaalisen analyysin evaluointi**

Molempien kokeiden tuloksina saatiin myös kartalle näkyviin alueet, joissa käyttäjät ovat keskustelleet aiheista. Analyysien tulokset ovat nähtävissä kuvassa 7 erivärisinä alueina. Ensimmäisen kokeen tuloksena saadut kommentit sijaitsivat kartalla melko erillään toisistaan. Alueet sijoituivat pääasiassa teiden risteyskohtiin, joihin liikenneaiheisten ongelmien olettaisikin usein sijoittuvan.

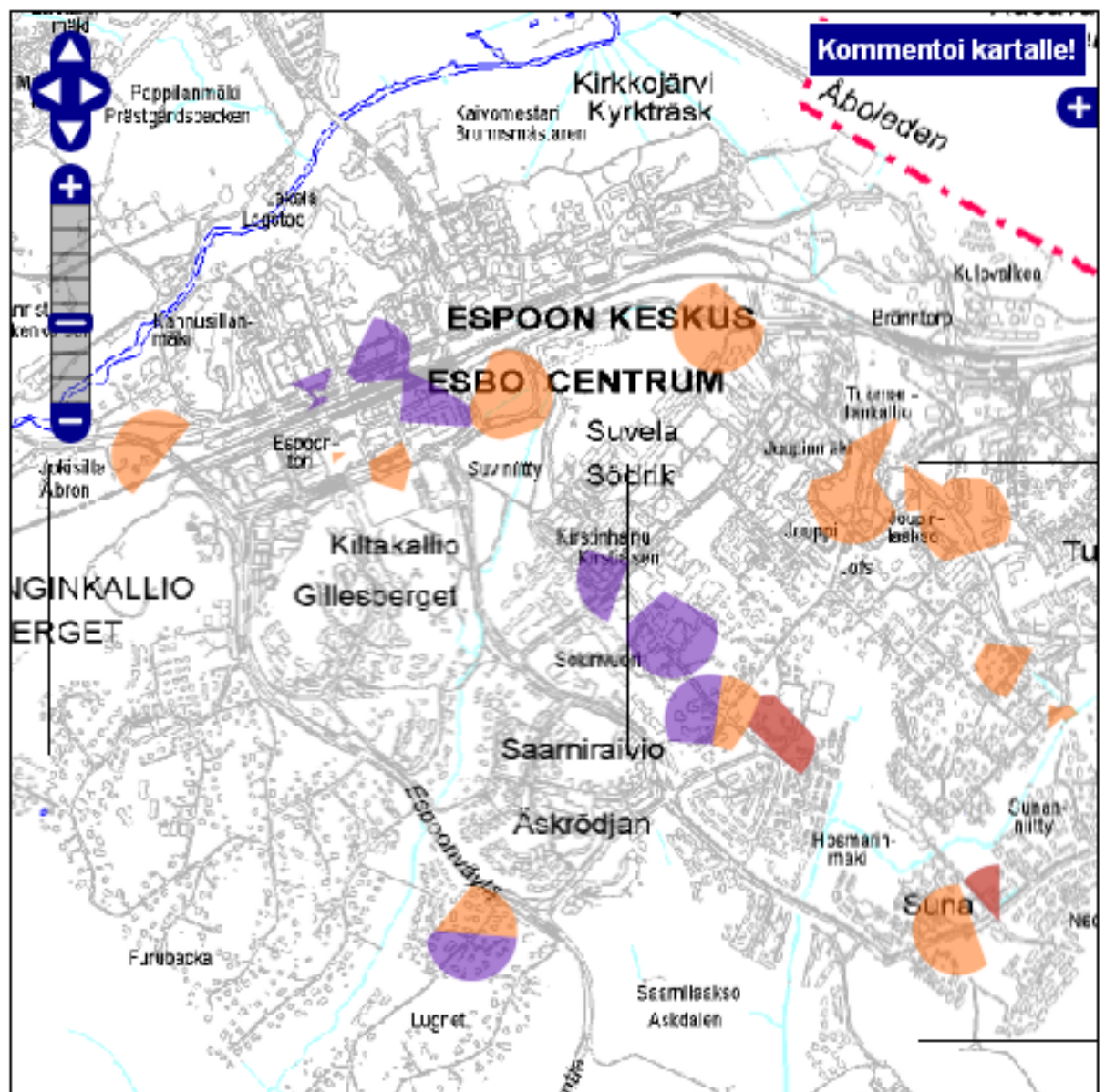
Toisen kokeen perusteella voitiin tutkia tarkemmin palvelun maantieteellisten ominaisuuksien toimintaa. Koska hakuterminä käytettiin paikan epämiellyttävyyttä kuvaavia sanoja, tulosten voitiin odottaa osuvan kartalla paikkoihin, jotka yleisesti koetaan epämiellyttävinä.

14:stä hakutuloksesta Espoon Keskusten juna-aseman alueelle osui 5 kommenttia. Asemasta noin kilometrin päässä Suvelassa sijaitsevan vain noin 500 metrin mittaisen Sokinsuontien varrelle osui 4 kommenttia. Loput 5 kommenttia sijaitsivat kauempana asemasta muodostaen yksittäisiä alueita. Tuloksen mukaan aseman seudun ja Suvelan pitäisi siis olla kommentoijien mielestä erityisen luotaantyöntäviä alueita.

Kehittämisfoorumin omalla paikkatietojärjestelmästä erillisellä keskustelupalstalla käydään paljon keskustelua Espoon Keskusten juna-aseman epäviihtyisyyteen liittyen (Rantanen 2008). Aseman seudun koetaan keskustelujen perusteella huonontavan koko alueen imagoa. Suvelaa puolestaan on lehdistössä luonnehdittu ”ongelmalähiöksi” (STT-IS 2007).

Espoon alueellisesta segregoitumisesta kertovassa Espoon Erilaiset Alueet - taustaselvityksessä vuodelta 2001 asuinalueen arvostuksen mittarina käytettiin asuntojen hintatasoa. Suvelan alueella asuntojen hinnat olivat Espoon alhaisimmat. Espoon voimakkain työttömyyskeskittymä oli Espoon keskuksen ja Suvelan alueella. Raportissa työttömyyden todetaan olevan segregoitumiseen liittyen merkittävä tekijä. (Espoon kaupunki 2001)

Kokeen tulosten perusteella analyysipalvelun voidaan siis todeta ainakin tietyissä tapauksissa tunnistavan erinomaisesti haettuja erityispiirteitä omaavia alueita.



Kuva 7, Analyysin tulos hakutermeillä ”pyörä” (oranssilla) ja ”likainen ankea ahdistava ruma hyi yäk ällöttävä epäsiisti” (violetilla) Espoon Keskukseen alueelta. Leikkaavat alueet punertavalla värillä.

## **6.2 Palvelun vaatimusten toteutuminen**

Palvelun toimintaa arvioidaan kappaleissa 3.3 ja 4.4 asetettujen vaatimusten valossa. Seuraavissa kappaleissa käsitellään kussakin ensin yleisesti vaatimuskategoriaan liittyviä piirteitä, minkä jälkeen vaatimusten täyttyminen käydään läpi kohta kerrallaan. Kategorian vaatimusten täyttyminen on esitetty taulukkomuodossa kunkin kappaleen lopussa.

### **6.2.1 Käyttäjien tarpeiden toteutuminen**

Käyttäjien moninaiset tarpeet kuvaavat parhaiten sitä, mitä analyysipalvelulta odotettiin. Erityisesti kaivattiin toimivaa työkalua laadullisen kommenttiaineiston käsittelyyn kaupunkisuunnittelijoiden työn helpottamiseksi.

**K1: Informatiivisuus.** Palvelun halutaan paljastavan jotakin todellista informaatiota siten, että sen perusteella voidaan tehdä johtopäätöksiä. Koekäytössä palvelun todettiin parantavan tapaa, jolla kommentteja käydään läpi. Etenkin ylläpitäjältä saattoi aiemmin helposti jäädä vanhat kommentit lukematta uudestaan, mutta analyysipalvelun tuloksia tarkastellessa aineistoa tulee käytyä läpi monipuolisemmin. Tilastollisen tekstianalyysin yhtenä tarkoituksena on nimenomaan jalostaa informaatiota helpommin tulkittavaan muotoon, josta on helpompi tehdä johtopäätöksiä. Tämä on myös yksi visualisoinnin päämääristä. Tämän työn puitteissa ei voitu kuitenkaan kattavasti testata palvelun vaikutuksia kaupunkisuunniteluun liittyvään päätöksentekoprosessiin.

**K2: Tiedon oikeellisuus.** Evaluointitesteissä palvelun todettiin kykenevän paljastamaan tekstiaineiston piilevän semanttisen rakenteen sekä osoittamaan tiettyjä ominaispiirteitä omaavia alueita. Toisaalta osa tuloksista oli myös virheellisiä. Yhdistettynä käyttäjän omaan harkintakykyyn ja mahdollisuuteen tehdä myös omia tulkintoja kommenttitekstien sisällöistä, palvelun avulla voidaan kuitenkin hankkia totuudenmukaista tietoa. Virhetulkinnat voi välttää suhtautumalla oikein analyysin tilastolliseen luonteeseen.

**K3: Visuaalinen esitystapa.** Analyysin tulos esitetään kartalla visuaalisessa muodossa

värillisinä alueina hyvän visualisointitavan periaatteita noudattaen. Koekäytössä analyysituloksen näkyminen kartalla todettiin hyödylliseksi ominaisuudeksi. Tuloksen paikantaminen on työtä, jonka käyttäjä tekisi itse tavalla tai toisella, jos se ei tapahtuisi automaattisesti.

**K4: Aineiston spatiaalisuuden huomioonottaminen.** Paikkatietoa hyödynnetään tekstianalyysin jälkeen suoritettavassa spatiaalisessa analyysissä, jossa alueiden rajausta suoritetaan kommenttien sijaintien mukaan suhteessa maastoon ja toisiinsa. Spatiaalisuus olisi mahdollista ottaa huomioon myös tekstianalyysissä, mutta sitä ei kokeiltu tämän työn puitteissa.

**K5: Aineiston kokonaisvaltainen käsittely.** Latentti semanttinen indeksointi on globaali siinä mielessä, että se käsittelee koko aineistoa kerrallaan. Mikä tahansa muutos missä tahansa kohdassa tekstiaineistoa voi periaatteessa muuttaa analyysin lopputulosta. Myös spatiaalisessa analyysissä käytetty Voronoi-diagrammi käsittelee koko aineistoa kerrallaan, tosin kaukana toisistaan olevat kommentit eivät vaikuta toisiinsa.

**K6: Ei rajoita muuta käyttöä.** Analyysipalvelun käyttöliittymä on sijoitettu kokonaisuudessaan oman välilehden alle siten, että paikallistietokartan muita ominaisuuksia käytettäessä näkyvissä on vain kyseisen välilehden otsikkokenttä. Palvelun olemassaolo ei myöskään kuormita palvelinta tai tietoliikenneyhteyttä havaittavissa määrin. Analyysin tuloksen saa palvelun käytön jälkeen helposti poistettua kartalta.

**K7: Helppokäyttöisyys.** Palvelu on käytettävyydeltään yksinkertainen ja intuitiivinen, suorastaan minimalistinen. Käyttäjää ei kuormiteta kysymyksillä tai vaihtoehtoilta. Hakutermiksi kelpuutetaan lähes mitä tahansa tekstiä, aina yksittäisestä sanasta kokonaiseen kommenttitekstiin. Myös taivutetut sanat tunnistetaan oikein, joissakin tapauksissa jopa väärin kirjoitetut. Kokemattoman käyttäjän voi kuitenkin olla vaikea arvata, mitä palvelun on tarkoitus tehdä, ja tarvitsee aluksi apua ainakin tulosten tulkinnassa. Käyttöliittymään liitettävän ohjeistuksen uskotaan auttavan asiaa. Paikkatietojärjestelmiin perehtyneen kaupunkisuunnittelijan omassa käytössä palvelun todettiin olevan helppo käyttää.

*Taulukko 11, Palvelun käyttäjävaatimusten toteutuminen*

K1 Informatiivisuus	Tavoitteet toteutuivat, mutta jatkotutkimusta kaivataan
K2 Tiedon oikeellisuus	Toteutui
K3 Visuaalinen esitystapa	Toteutui
K4 Aineiston spatiaalisuuden huomioon ottaminen	Toteutui
K5 Aineiston kokonaisvaltainen käsittely	Toteutui
K6 Ei rajoita muuta käyttöä	Toteutui
K7 Helppokäyttöisyys	Ohjeistusta kaivataan

## 6.2.2 Yhteensopivuus Kehittämisfoorumin kanssa

Koska palvelu toteutettiin osana Kehittämisfoorumia, on tärkeää ettei muutos vähennä järjestelmän muiden ominaisuuksien arvoa.

**Y1: Alueen kehityksen tukeminen.** Palvelun avulla voidaan helpommin tuoda esiin kommenttien sisältämää paikallistietämystä. Paikallistietämystä voidaan hyödyntää kaupunkisuunnittelussa, jonka yleisenä päämääränä on suunniteltavan alueen kehittäminen.

**Y2: Avoimuus ja läpinäkyvyys.** Tekstianalyysimenetelmät ovat maallikoille melko tuntemattomia, joten ei voida olettaa, että kaikki käyttäjät tietäisivät miten analyysi tarkalleen tapahtuu. Maallikkokäyttäjät näkevät tulokset kuitenkin samalla tavalla kuin asiantuntijatkin. Palvelun avoimuus tuo kommenttien käsittelyprosessia läpinäkyvämmäksi, kun kansalaiset voivat itse kokeilla, miten heidän kommentointiaan voidaan käyttää kaupunkisuunnittelun osana.

**Y3: Puolueettomuus.** Palvelun puolueettomuus on seurausta käytetystä ohjaamattomasta menetelmästä, joka suhtautuu kommentteihin tasapuolisesti. Palvelun toimintaan ei ole ohjelmoitu erityiskäsittelyä tietyn tyyppisille kommentteille, mikä voisi vaikuttaa palvelun puolueellisuuteen.

**Y4: Ajankohtaisuus.** Analyysipalvelu ei voi sinänsä vaikuttaa kommenttien puheenaiheiden ajankohtaisuuteen, mutta se huolehtii ajankohtaisuudesta omalta osaltaan käyttämällä aina tietokannassa olevaa uusinta aineistoa. Jos kommenttiaineisto muuttuu esimerkiksi uuden kommentin vuoksi, koko aiheavaruus viritetään seuraavalla analyysikerralla uudestaan.

**Y5: Ylläpidon edullisuus ja paikallisuus.** Analyysipalvelun lisäys ei aiheuttanut suuria muutoksia ylläpidon työmäärään. Ylläpito voidaan edelleen toteuttaa Arkkitehtuurin laitoksella.

**Y6: Alueen imagon parantaminen.** Kuten toisessa evaluointikokeessa havaittiin, analyysipalvelulla on mahdollista kartoittaa alueen ongelmapaikkoja. Kaupunkisuunnittelussa, kuten suunnittelussa yleensäkin, ongelmien paikantaminen on tärkeätä ratkaisun kehittämiseksi. Vaikutuksia alueen imagon kohentumiseen ei voitu kuitenkaan työssä todeta.

**Y7: Käytettävyydeltään yhdenmukainen muun järjestelmän kanssa.** Koekäytössä havaittiin monia yksittäisiä käytettävyydspuutteita, joista useimmat liittyivät laajemmin koko järjestelmän käytettävyyteen, eikä niitä siten voitu tämän työn puitteissa ratkaista. Yleisesti käytettävyyden todettiin olevan Kehittämisfoorumin tason mukainen. Analyysipalvelu toimii paikkatietojärjestelmän luontevana osana samaan tapaan kuin järjestelmän muutkin toiminnot.

**Y8: Kopioitavuus.** Palvelu voidaan suhteellisen helposti kopioida toiseen samoilla teknologioilla toteutettuun kehittämisfoorumiin sen modulaarisuudesta johtuen. Koska käytetty menetelmä on ohjaamaton, sitä voidaan käyttää muissa ympäristöissä ilman, että aineiston erityispiirteitä tarvitsee ottaa huomioon.



*Taulukko 12, Palvelun yhteensopivuusvaatimusten toteutuminen*

Y1 Alueen kehityksen tukeminen	Toteutui
Y2 Avoimuus ja läpinäkyvyys	Toteutui
Y3 Puolueettomuus	Toteutui
Y4 Ajankohtaisuus	Toteutui
Y5 Ylläpidon edullisuus ja paikallisuus	Toteutui
Y6 Alueen imagon parantaminen	Tavoitteet toteutuivat, mutta jatkotutkimusta kaivataan
Y7 Käytettävyydeltään yhdenmukainen	Toteutui
Y8 Kopioitavuus	Toteutui

### **6.2.3 Osallistavuuden edistäminen**

Demokratiaa, kansalaisten voimaantumista ja osallistamisen tasoa on vaikea mitata (Zimmerman & Rappaport 1988) (Saward 1994). Sen vuoksi on vaikea myös todentaa yksittäisen muutoksen vaikutusta osallistavuuteen. Osallistavuuden kehitystä voidaan kuitenkin arvioida kirjallisuuden perusteella. McCallin ja Arnsteinin osallistavuusasteikoille sijoitettuna Kehittämisfoorumin paikallistietokartta osallistavana paikkatietojärjestelmänä tarjoaa konsultointitason palveluita antamalla osallisille mahdollisuuden aktiivisesti osoittaa uusia ongelmakohtia. Kehitetty analyysipalvelu ei nosta osallistavuutta osallisten päätöksentekoon mukaan ottamisen tai Arnsteinin määrittelemän tyynnyttelyn tasolle. Asteikko on kuitenkin niin karkea, että muutoksella voi olla vaikutusta osallistavuuteen ilmeisesti, että se nostaa osallistamisen tasoa

kokonaisella pykälällä ylöspäin. Kehitetty palvelu parantaa kansalaisten konsultointiin käytettävää menettelyä helpottamalla tiedonkeruuta ja vähentämällä epätietoisuutta siitä, miten kerättyä tietoa käsitellään. Siten palvelu voidaan nähdä edistysaskeleen sijaan myös saavutetun aseman vahvistamisena.

**O1: Tasavertaisuus.** Käytetty menetelmä käsittelee kaikkia kommentteja tasapuolisesti sijoittamalla ne aiheavaruuteen niiden puheenaiheiden mukaan. Kaikkien kommenttien oletetaan kertovan omasta aiheestaan yhtä vahvasti. Kommentteja ei myöskään luokitella sen mukaan, kuinka moni niistä kertoo samasta aiheesta, joskin tällä on vaikutusta aihealueiden muodostumiseen. Kommentin päätyminen hakutulokseen riippuu ainoastaan hakutermin ja kommentin puheenaiheen yhtäläisyydestä.

**O2: Saavutettavuus.** Toinen osallistavuuden perusvaatimuksista, saavutettavuus, kärsii yleensä eniten suunnittelun aika- ja paikkasidonnaisuudesta. Kehitetty palvelu toimii avoimessa julkisessa paikkatietojärjestelmässä. Internet ei sinänsä takaa palvelun saavutettavuutta kaikille, mutta sen on todettu olevan paras alusta osallistaville palveluille. Siten palvelu on käytössä ”missä ja milloin vain” siinä määrin kuin tämä nykypäivänä on ylipäänsä mahdollista, mobiilisovelluksia lukuun ottamatta.

*Taulukko 13, Palvelun osallistavuusvaatimusten toteutuminen*

O1 Tasavertaisuus	Toteutui
O2 Saavutettavuus	Toteutui

## 6.2.4 Teknisten vaatimusten toteutuminen

Palvelun toiminnan perustana toimii tekniikka, jolla analyysin tulos tuotetaan kerta toisensa jälkeen käyttäjän selainikkunaan. Tekniset näkökohdat ovat usein käyttäjälle huomaamattomia niin kauan kuin palvelu toimii. Kappaleessa 3.3.4 listattiin vaatimuksia,

joilla mahdollistetaan palvelun sujuva toteutus, käyttö, ylläpito ja jatkokehitys.

**T1: Muutoksen yksinkertaisuus käyttäjän kannalta.** Liian suuret muutokset järjestelmän toiminnassa voivat aiheuttaa käyttäjän turhautumista, kun hänen pitää jokaisen muutoksen jälkeen opetella käyttö uudestaan. Kehitetty palvelu ei juurikaan vaikuta Kehittämisfoorumin muuhun toimintaan eikä välttämättä muuta yksittäisen käyttäjän tapaa käyttää sivustoa. Palvelun ominaisuuksien opettelu voi kuitenkin viedä oman aikansa ja sitä käsiteltiin helppokäyttöisyysvaatimuksessa K7.

**T2: Modulaarisuus.** Palvelun keskeisenä osana toimii Drupal-moduli, joka nimensä mukaisesti liittyy sisällönhallintajärjestelmään modulaarisesti. Palvelun muu toiminta kommunikoi suoraan vain modulin kanssa, ja palvelu voidaan poistaa käytöstä sammuttamalla moduli sisällönhallintajärjestelmän ylläpitonäkymästä.

**T3: Toteuttamiskelpoisuus resurssien puitteissa.** Työn toteutuksella oli melko tiukka aikataulu, ja toimintaa saneli usein samassa yhteydessä tapahtuva järjestelmän alustanvaihdos. Aikataulussa pysymistä vaikeutti osaltaan työn laajuus ja tarvittavien ohjelmointitaitojen hankkimiseen kuluva aika. Työ saatiin kuitenkin toteutettua ja testattua aikataulussaan.

**T4: Nopea vasteaika.** Analyysipalvelun toimintaa saatiin nopeutettua huomattavasti tallentamalla pääkomponenttianalyysin tuottama aiheavaruus tiedostoon ja käyttämällä sitä myöhemmillä analyysikerroilla. Analysointi kestää noin 30-60 sekuntia riippuen käytetystä tietoliikenneyhteydestä ja asiakaspääteestä. Aika on pitempi kuin paikallistietokartan muiden toimintojen vaatima aika, mutta ei kuitenkaan niin pitkä, että käyttäjä ehtisi turhautua analyysin aikana tai joutuisi poistumaan päätteensä äärestä odottamaan sen valmistumista. Tulosta odotellessa voi samalla käyttää paikkatietojärjestelmän muita toimintoja tai tarkastella aiemmin haettuja analyysituloksia.

**T5: Aineiston synkronointi:** Informaation redundanssin ja aineiston synkronoinnin ongelmat vältetään, koska aineistoa säilytetään vain yhdessä paikassa. Tällöin järjestelmän eri osat eivät voi pitää omia paikallisia tietokantoja aineiston tilasta.

Ainoana poikkeuksena on erilliseen tiedostoon tallennettava aiheavaruus, joka lasketaan uudestaan aina kun aineiston havaitaan muuttuneen.

**T6: Alustariippumattomuus:** Palvelun toimintaa kokeiltiin muutamilla eri päätteillä sekä yleisimmillä Internet-selaimilla. Toiminnassa ei havaittu eroja ei kokoonpanoilla.

*Taulukko 14, Palvelun teknisten vaatimusten toteutuminen*

T1 Muutoksen yksinkertaisuus käyttäjän kannalta	Toteutui
T2 Modulaarisuus	Toteutui
T3 Toteuttamiskelpoisuus resurssien puitteissa	Toteutui
T4 Nopea vasteaika	Toteutui
T5 Aineiston synkronointi	Toteutui
T6 Alustariippumattomuus	Toteutui

### 6.2.5 Palvelun menestymisvaatimusten toteutuminen

Työn suunnittelussa ja toteutuksessa pyrittiin hyödyntämään aiemmista vastaavanlaisista palveluista opittuja tietoja. Moniin listattuihin asioihin ei oltaisi voitu vaikuttaa, vaan työ toteutettiin niissä olosuhteissa kuin se oli mahdollista.

**M1: Toimivaltainen kehitysyhteisö.** Kehitystyö tapahtui korkeakoulussa, jolla on hyvät yhteydet myös yksityiseen sektoriin. Työn rinnalla tapahtuneen alustanvaihdoksen toteutti yksityinen yritys, joka toimi tukena myös analyysipalvelun kehittämisessä.

**M2: Julkisen hallinnon tuki.** Kehitystyö tehtiin julkisella rahoituksella. Järjestelmän omistaa Espoon kaupunki.

**M3: Roolien selkeys.** Työllä oli selkeästi erilliset roolit: toteuttaja, ohjaajat, valvoja, koekäyttäjä, ylläpitäjät.

**M4: Osa laajempaa kokonaisuutta.** Teknisesti palvelu toteutettiin osana osallistavaa paikkatietojärjestelmää. Kehitystyö on samalla myös osa Arkkitehtuurin laitoksen tutkimustoimintaa.

**M5: Toteutuksen iteratiivisuus.** Palvelu toteutettiin proof-of-concept -hengessä ja vaikka se otetaan hyötykäyttöön, se on prototyyppi. Palvelu ei ollut toteutuksen aikana varsinaisessa käytössä. Myöhemmin jatkokehityksessä voidaan kuitenkin hyödyntää saatuja kokemuksia.

**M6: Monitavoitteisuus.** Työn tarkoituksena oli tutkia tekstianalyysin hyödyntämistä kaupunkisuunnittelussa osallistavan paikkatietojärjestelmän tapauksessa. Lisäksi tavoitteena oli kehittää kaupunkisuunnitteluun toimiva apuväline sekä parantaa yleisesti kansalaisten osallistavuutta Espoon keskuksen alueella.

*Taulukko 15, Palvelun yleisten menestymisvaatimusten toteutuminen*

M1 Toimivaltainen kehitysyhteisö	Toteutui
M2 Julkisen hallinnon tuki	Toteutui
M3 Roolien selkeys	Toteutui
M4 Osa laajempaa kokonaisuutta	Toteutui
M5 Toteutuksen iteratiivisuus	Toteutuminen riippuu jatkokehityksestä
M6 Monitavoitteisuus	Toteutui

## 6.2.6 Aineiston vaatimukset

Käytetty menetelmä asetti aineistolle seuraavat vaatimukset.

**A1: Sopivan kokoinen aineisto.** Latentti semanttinen indeksointi vaatii toimiakseen yleensä muutaman sadan sanan aineiston. Käytetty 208:n kommentin aineisto oli pienehkö, mutta tuloksista päätellen se oli kuitenkin riittävän suuri. Mahdollisesti esimerkiksi puolet pienemmällä aineistolla tulokset eivät olisi olleet tyydyttäviä. Käytetty aineisto ei myöskään ollut niin suuri, että sen käsittely olisi ollut haitallisen hidasta.

**A2: Riittävän pitkät kommentit.** Kommenttitekstien pituudella on samankaltainen vaikutus menetelmän toimintaan kuin kommenttien määrällä. Nyt osa kommenteista oli vain parin sanan mittaisia, mikä vaikeutti niiden aiheen tunnistamista. Aineistosta löytyi myös muutama kommentti, joiden sanoista yhtäkään ei löytynyt muista kommenteista, mutta nämä olivat sisällöstä päätellen kaikki testikommentteja. Pisimmissä kommenteissa puolestaan puhuttiin usein monesta eri puheenaiheesta, mikä oli käytetyn menetelmän kannalta ongelmallista, koska aiheavaruuden perusteella kommentin tulkitaan kuitenkin kertovan vain yhdestä aiheesta. Menetelmä sijoittaa kommentin tällaisessa tapauksessa aiheiden puoliväliin, mikä tosin on usein riittävän hyvä ratkaisu. Useimmat kommenteista olivat kuitenkin tuloksista päätellen sopivan mittaisia ja yhteisiä sanoja löytyi eri kommenteista riittävästi.

**A3: Aineiston eheys.** Aineiston laatu vaikuttaa myös sen analysointitulokseen. Erityisenä ongelmana latentin semanttisen indeksoinnin kohdalla havaittiin aineistossa moneen kertaan olevat samat kommentit. Koska menetelmä tekee päätelmiä sanojen yhtäläisyyksistä niiden yhteisesiintyvyyden mukaan, duplikaattikommentit vääristävät aiheavaruutta. Ongelma saatiin korjattua esikäsittelemällä aineisto siten, että kommenttien kopiot jätetään kokonaan käyttämättä aiheavaruuden virityksessä.

*Taulukko 16, Käytetyn aineiston vaatimusten toteutuminen*

A1 Sopivan kokoinen aineisto	Toteutui
A2 Riittävän pitkät kommentit	Toteutui
A3 Aineiston eheys	Toteutui

### **6.3 Palvelun arviointi kokonaisuutena**

Palvelua arvioitiin evaluoinnin ja vapaamuotoisen koekäytön menetelmillä sekä vertailemalla saatuja tuloksia ennalta asetettuihin vaatimuksiin. Palvelun toiminnasta heräsi myös ajatuksia, jotka eivät suoraan liity vaatimuksiin.

Palvelu suunniteltiin sekä ammattilais- että maallikkokäyttöön. Koekäyttäjä arvioi palvelun käytön kuitenkin sen verran hankalaksi, että maallikko tarvitsisi ohjeistusta tai jopa perehdytystä sen käytössä. Tämä on osallistavuuden kannalta huono asia, koska perehdytystä ei voida järjestää tasapuolisesti kaikille alueen toimijoille. On kuitenkin mahdollista, että palveluun liitettävällä hyvällä ohjeistuksella saadaan kaikki käyttäjät oppimaan analyysipalvelun käytön.

Analyysipalvelun, kuten minkä tahansa apuvälineen, käyttäminen on taito, jossa voi kehittyä harjoittelemalla. Koekäyttäjä oppi käytön aikana pikkuhiljaa käyttämään yhä toimivampia hakutermejä. Kansalaisten osallistamisen kannalta on positiivista, että kuka tahansa voi halutessaan opetella käyttämään palvelua. Toisaalta olisi kuitenkin parempi, jos käyttöä ei tarvitsisi lainkaan opetella, vaan kaikilla olisi suoraan samat mahdollisuudet palvelun käyttöön, perehtyneisyydestä riippumatta. Koekäyttäjänä toimiva kaupunkisuunnittelija totesikin, että kansalaisia ei pidä tasapäistää, vaan heilläkin voi olla hyvinkin erilaiset valmiudet uudenlaisen toiminnon opetteluun.

Tekstianalyysi toimii yllättävän hyvin käytetyn menetelmän yksinkertaisuudesta huolimatta – tai siitä johtuen. Vaikka menetelmän toteutukseen käytettiin paljon aikaa, ratkaisut joihin päädyttiin olivat enimmäkseen kirjallisuudesta tuttuja latentin semanttisen indeksoinnin perusominaisuuksia. Menetelmään ei jouduttu lisäämään

minkäänlaisia erikoiskäsittelyjä erikoistapauksia varten, vaan koko aineistoa käsitellään samalla tavalla sen sisällön mukaan.

Menetelmä toimii hyvin erilaisilla hakutavoilla. Hakuterminä voidaan käyttää yhtä tai useampaa sanaa. Sanat voivat muodostaa lauseen tai olla lista synonyymeistä. Sanojen ei välttämättä tarvitse edes liittyä samaan aiheeseen, vaan voidaan hakea ikään kuin sanojen väliin sijoittuvaa aihetta. Lisäksi käyttäjä voi käyttää aiemmassa haussa löytynyttä kommenttitekstiä uutena hakuterminä, jolloin hän löytää kyseiseen kommenttiin eniten liittyvät muut kommentit. Käyttäjän ei myöskään tarvitse miettiä hakusanojen taivutusmuotoja. Nämä kaikki ominaisuudet ovat emergenttiä seurausta pelkästään latentista semanttisesta indeksoinnista, kommenttien ja hakutermin skaalauksesta samanmittaisiksi, ja sanaston yksinkertaisesta esikäsittelystä.

Koekäytössä havaittiin erilaisiin hakutapoihin liittyen, että useammasta sanasta muodostuvalla hakutermillä saadaan usein osuvampi tulos kuin yksittäisellä sanalla. Samanaiheisia sanoja hakuterminä käytettäessä tämä selittyy sillä, että hakutermin tunnistetaan tilastollisessa mielessä tarkemmin, kun useampi sana vie aihevektoria käyttäjän haluamaan suuntaan. Jokaisen yksittäisen sanan sijainnissa aiheavaruudessa on tilastollista virhettä, joka tasoittuu keskiarvoistumisen myötä useampaa sanaa käytettäessä. Tämän perusteella monien samanaiheisten sanojen käyttö on suositeltava hakutapa. Sanoja voi etsiä aineistosta iteratiivisesti käyttämällä hakuterminä ensin yhtä sanaa, ja sitten lisäämällä siihen sopivia sanoja löytyvien kommenttien joukosta.

Puheenaiheiden visualisoinnissa käytetty menetelmä todettiin palvelun käytettävyyden kannalta hyväksi, koska sen avulla pystytään paikantamaan hakutulokseen kuuluvat kommentit. Alueiden yhtenäisyys helpottaa tiedon luettavuutta tapauksissa, joissa hakutulokset sijaitsevat vierekkäin.

Kokeellisessa sovelluksessa toteutetun epämääräisen informaation visualisoinnin onnistumista ei voida määrittää suoraan kommunikaation tehostumisen mittareilla, vaan on tutkittava sen vaikutuksia esimerkiksi päätöksentekoon, hahmontunnistukseen, hypoteesien muodostamiseen tai käytössä oleviin käytäntöihin (MacEachren 1992). Alueiden maantieteellisen muodon mielekkyyttä ei voitu koekäytössä verifioida, vaan



sen selvittämiseksi olisi tutkittava maaston todellisia piirteitä tai ihmisten mielikuvia siitä suhteessa visualisoinnin tulokseen. Alueiden rajojen muodostumisen kriteerinä ei kuitenkaan ollut kuvata tarkasti maaston piirteitä, vaan auttaa käyttäjää muodostamaan selkeä mielikuva puheenaiheiden sijoittumisesta, missä se onnistuu testien perusteella hyvin.

## **6.4 Toteutetun palvelun jatkokehitys**

Työn kokeellisen osan valmistuminen ei tarkoita sitä, että kehitetty palvelu olisi valmis. Työn arvioinnin myötä saadaan paljon viitteitä siitä, miten palvelua voitaisiin kehittää edelleen. Myös työn suorituksen aikana keksittiin paljon ideoita, jotka työn aikataulusta ja rajauksista johtuen jätettiin toteuttamatta. Tässä kappaleessa on esitetty niistä tärkeimmät.

### **6.4.1 Tekstianalyysin vaihtoehtoisia toimintatapoja**

Työssä käytettiin dimension pienennykseen vain pääkomponenttianalyysia. Yhtä lailla olisi voitu käyttää myös riippumattomien komponenttien analyysia tai itseorganisoituvaa karttaa. Jatkokehityksessä olisi syytä kokeilla, päästäisiinkö näitä käyttämällä parempiin tuloksiin.

Säilytettävien pääkomponenttien määrä on parametri, jolle täytyy yleensä etsiä sopiva arvo kokeilemalla analyysin tuloksia eri arvoilla. Toinen vaihtoehto olisi käyttää hakutermitä riippuvaa määrää pääkomponentteja. Pääkomponentteja voidaan käyttää esimerkiksi niin monta kuin tarvitaan, että hakutermien aihevektori saadaan vähintään tietyn mittaiseksi. Aihevektorista tulee sitä pitempi, mitä enemmän pääkomponentteja säilytetään, koska jokainen uusi käytettävä ulottuvuus voi vain kasvattaa vektorin pituutta.

Menettely voidaan mieltää siten, että hakutermien aihetta tutkitaan niin moniulotteisessa avaruudessa, kuin on tarpeen aiheen tunnistamiseksi. Aihevektorin ei kuitenkaan kannata olla liian pitkä, koska halutaan jättää tilaa myös latentin semanttisen indeksoinnin tarvitsemalle epämääräisyydelle, jonka avulla sanat voivat yhdistyä

puheenaiheiksi.

Kommenttien otsikko- ja sisältötekstistä löytyviä sanoja käytetään keskenään samalla tavalla. Olisi mahdollista myös painottaa niistä löytyviä sanoja sen mukaan, miten paljon niillä uskotaan olevan kontribuutiota kommentin todellisen puheenaiheen muodostumiseen. Otsikosta löytyville sanoille voitaisiin käyttää esimerkiksi kaksinkertaista painokerrointa sisältötekstistä löytyviin verrattuna.

#### **6.4.2 Vaihtoehtoja käytettävän aineiston suhteen**

Aiheavaruuden virittämiseen käytettävän aineiston määrää voi tarvittaessa kasvattaa käyttämällä ulkopuolista tekstiaineistoa, ns. *korpusta*, jonka ei tarvitse olla paikkatietoa. Jos avaruus viritetään suuremman tekstiaineiston perusteella, sanojen välisiä yhteyksiä löytyisi paljon enemmän, mikä tekisi aiheavaruudesta tilastollisessa mielessä tarkemman. Korpuksen pitäisi kuitenkin kertoa samasta aiheesta ja samalla kirjoitustyyllillä kuin kommenttiaineiston, jotta aiheet muodostuisivat kehitettävää aluetta kuvaavalla tavalla. Eräs mahdollisuus olisi käyttää korpuksena Kehittämisfoorumin keskustelupalstan kommentteja.

Tekstianalyysissa käytetty menetelmä ei tunnista kaikkia kommenttiaineiston sanojen välisiä yhteyksiä oikein. Käyttäjille voitaisiin tarjota mahdollisuus yhdistellä sanojen merkityksiä itse, esimerkiksi nimeämällä kerrallaan kaksi toisiinsa liittyvää sanaa. Vastaavasti myös kokonaisia kommentteja voisi linkittää toisiinsa liittyviksi. Palvelu voisi kerätä yhteyksistä tietokantaa, joka täydentyisi pikkuhiljaa käytön myötä. Tällöin käyttäjillä olisi kuitenkin mahdollisuus manipuloida tulosta eikä analyysia tehtäisi enää pelkästään käytettävän aineiston perusteella, mikä heikentäisi sen läpinäkyvyyteen liittyvää luotettavuutta.

Espoon Keskusten Kehittämisfoorumin kohdalla on ristiriitaisia käsityksiä siitä, pitäisikö keskustelu alueen kehityksestä ensisijaisesti käydä paikallistietokartalla vai sivuston erillisellä keskustelupalstalla. Yksittäiseen karttakommenttiin liittyvä jatkokeskustelu käydään usein keskustelupalstalla.

Toteutetun palvelun toiminnan ja siitä saatavan hyödyn kannalta olisi parasta, jos keskustelu käytäisiin kokonaisuudessaan paikallistietokartalla. Tällöin aineistoa olisi enemmän. Yhteen kommenttiin liittyvä kommentointi olisi hyödyllistä saada osaksi kommentin sisältöä, jolloin koko keskustelua voitaisiin käsitellä yhtenä tekstinäytteenä kommenttirykelmän sijaan. Tällöin sen puheenaihe tunnistettaisiin tarkemmin, koska sanoja olisi enemmän. Lisäksi spatiaalinen visualisointi esittäisi alueen joko kokonaan aiheeseen liittyvänä tai ei-liittyvänä sen sijaan, että alueen rajat kulkisivat yksittäisten väitteiden välissä maastoon liittymättömällä tavalla.

### **6.4.3 Alueiden rajaamisen kehitys ryhmittelyn avulla**

Puheenaiheiden visualisointitapa herätti koekäytössä hämmästyttäviä tilanteita, joissa eri aiheista kertovat kommentit sijaitsivat aivan vierekkäin. Tällöin puheenaiheiden rajat näyttävät kulkevan kommenttien lävitse ja rajan suunta riippuu voimakkaasti siitä, mihin pisteisiin kommentit on satuttu tarkalleen sijoittamaan. Monissa tilanteissa vaikutti siltä, että kommentit oli tarkoitettu täsmälleen samaan pisteeseen mutta karttaliittymän epätarkkuudesta tai muusta syystä johtuen ne eivät ole aivan samassa kohdassa.

Tällaisissa tilanteissa voitaisiin alue määrittää eri tavalla. Koska kommentit ovat sijoittuneet samaan paikkaan, niiden puheenaiheena on ainakin jossakin määrin kyseinen maastonkohta. Kommentteille voidaan käyttää yhteistä Voronoi-aluetta, jonka aihe päätetään molempien kommenttitekstien sisällön mukaan. Aiheet voidaan yhdistää esimerkiksi käyttämällä niiden aihevektorien keskiarvoa. Yhteisenä Voronoi-alueena voidaan käyttää alueiden unionia, tai sen voidaan antaa muodostua itsestään poistamalla toinen pisteistä ennen Voronoi-diagrammin muodostusta.

Aina ei voida kuitenkaan olettaa, että samassa kohdassa sijaitsevat kommentit kertoisivat täysin samasta aiheesta, joten mainitun ominaisuuden toimintaa pitäisi arvioida testitulosten perusteella ennen sen käyttöönottoa.

### **6.4.4 Spatiaalinen tekstianalyysi**

Työn tavoitteena oleva tekstianalyysin ja maantieteellisen visualisoinnin yhdistäminen

toteutui. Spatiaalisuus voitaisiin kuitenkin tuoda myös osaksi tekstianalyysia käyttämällä spatiaalimuuttujia sanojen rinnalla pääkomponenttianalyysissa. Yksinkertaisimmillaan tämä voisi tarkoittaa kahden uuden, kommenttien X- ja Y-koordinaatit esittävän, ulottuvuuden lisäämistä sana-avaruuteen ennen pääkomponenttianalyysia. Kyseisten ulottuvuuksien skaalauksella voitaisiin vaikuttaa suuresti siihen, kuinka suurelta osin kommentin sijoittuminen aiheavaruudessa tapahtuisi kommentin fyysisen sijainnin ja toisaalta sen tekstisisällön mukaan.

Spatiaalimuuttujien lisäämisen vaikutus olisi se, että fyysisesti toisiaan lähellä olevien kommenttien sanat sijoittuisivat aiheavaruudessa lähemmäs toisiaan kuin muuten. Tällöin esimerkiksi yhden tienristeyksen kohdalla ryppäässä olevien kommenttien oletettaisiin kertovan ainakin osittain samasta aiheesta, nimittäin kyseisestä risteyksestä. Ilman spatiaalimuuttujia kommenttien puheenaiheiden yhteys kyseiseen risteykseen tapahtuu vain niiden tekstisisällön mukaan. Toisin sanoen tekstissä pitää tavalla tai toisella viitata kyseiseen paikkaan. Koordinaattien huomioiminen auttaisi myös sijoittamaan aiheavaruuteen sanoja, jotka löytyvät vain yhdestä kommentista, koska koordinaattien avulla sanat voidaan yhdistää yksittäisen kommentin lisäksi myös paikkaan.

Ominaisuus saattaisi parantaa myös palvelun visualisointiominaisuuksia tekemällä alueista yhtenäisempiä. Jos kartalla vierekkäin olevat kommentit sijaitsevat aiheavaruudessa lähellä toisiaan, ne päätyvät myös helpommin saman hakutermiin tuloksiksi. Koordinaateilla ei saisi kuitenkaan olla niin suurta vaikutusta, että hakutulokseen otettaisiin mukaan väärästä aiheesta kertovia kommentteja, jotka ovat aiheeseen kuuluvien kommenttien läheisyydessä. Sen sijaan ominaisuuden pitäisi olla hienovarainen säätötoimenpide, jolla parannetaan aiheen tunnistamista rajatapauksissa.

## **6.5 Palvelukokonaisuuden muiden palvelujen toteuttaminen**

Toteutettu palvelu oli vain yksi suunnitellun palvelukokonaisuuden ohjelmista. Toteutuksen tarkoituksena oli osoittaa, että tämän tyyppisiä palveluita voidaan toteuttaa ja luoda pohjaa jatkokehitykselle.

Kuten kappaleessa 5.4 todettiin, latenttia semanttista indeksointia ei toteutettu erillisenä itsenäisenä ohjelmana, vaan se toimii osana analyysin suorittavaa Perl-ohjelmaa.

Palvelukokonaisuuden toteuttamiseksi se olisi kuitenkin syytä ohjelmoida itsenäiseksi ohjelmaksi, jotta samaa toimintaa ei tarvitsisi ohjelmoida jokaiseen palveluun erikseen. Lisäksi tekstianalyysin erillisuus parantaa käytön tehokkuutta, kun aiheavaruus on aina valmiiksi laskettuna erillisessä tiedostossa myös aineiston muuttumisen jälkeen.

Kappaleessa 4.1.1 esitetty valitun karttapisteen puheenaiheet esittävä palvelu vaikuttaa erittäin toteutuskelpoiselta. Uutta tekniikkaa toteutettuun palveluun nähden on käyttäjän valitseman karttapisteen koordinaattien selvittäminen, jotta voidaan laskea pisteen etäisyys kommentteihin. Tämä onnistunee OpenLayersin omilla funktioilla.

Kappaleen 4.1.2 samankaltaisten kommenttien tunnistus voidaan toteuttaa puheenaiheen etsintänä käyttämällä valittua kommenttia suoraan hakuterminä.

Kappaleen 4.1.3 klusterointiominaisuuden toteuttaminen vaatii enemmän perehtymistä klusterointimenetelmiin. Lisäksi ongelmaksi voi koitua se, että OpenLayersista ei löydy valmista funktiota, jolla tekstiä, kuten puheenaiheita, voitaisiin esittää kartalla. Jotta alueista saataisiin mahdollisimman yhtenäisiä ja selkeitä, aiheavaruuden muodostamiseen on syytä kokeilla kappaleessa 6.4.4 esitettyä spatiaalista tekstianalyysia.

Eri palvelut voidaan sijoittaa paikallistietokarttanäkymässä esimerkiksi kukin omalle välilehdelle, ja niille voidaan suunnitella erilaiset käyttöliittymät. Eri analyysipalveluiden tulokset voidaan säilyttää kartalla näkyvissä siten, että käyttäjä voi helposti yhdistellä niistä saamiaan tietoja.

Puheenaiheita kartalla visualisoivaa palvelua ei valittu toteutettavaksi sen helppouden takia, vaan siitä uskottiin saatavan suurinta suoraa hyötyä kaupunkisuunnittelun ja osallistamisen tarpeisiin. Tulosten kannustamana voidaan todeta, että vastaavatyypisiä palveluita voitaisiin todennäköisesti kehittää lisääkin.

## 7 Johtopäätökset

Verkkopalvelu, jonka avulla osalliset voivat ilmaista näkemyksiään alueensa kehittämisestä, on teknisesti helppo toteuttaa. Palvelusta ei kuitenkaan ole hyötyä päätöksenteon kannalta, jollei tietoa pystytä keräämään ja käsittelemään riittävän tasapuolisesti ja jollei sitä pystytä riittävän tehokkaasti välittämään päätöksentekijöille. Suurimmat haasteet osallistavan GIS:n toteuttamisessa liittyvätkin tapaan, jolla tietoa kerätään ja esitetään.

Espoon Keskusten Kehittämisfoorumi kerää paikallisilta toimijoilta kommentteja tasapuolisella ja avoimella tavalla. Tietoa ei kuitenkaan jalosteta, vaan se esitetään sellaisenaan paikkatietojärjestelmän kartalla yksittäisinä pisteinä. Tässä diplomityössä kehitetty palvelu tunnistaa kommenttien puheenaiheet ja visualisoi käyttäjän etsimät puheenaiheet kartalle alueina. Näin käyttäjä saa paremman kuvan aiheen laajuudesta ja spatiaalisesta luonteesta ilman, että hänen tarvitsee käydä läpi kaikkien kommenttien sisältöjä läpi.

Aalto-yliopiston teknillisen korkeakoulun Arkkitehtuurin Laitoksen havaittiin olevan tämän tyyppisen palvelun kehittämiseen erittäin sopiva kehitysyhteisö, jossa työn tavoitteet voitiin pitää kaikin puolin mielekkäinä ja toisiaan tukevinä. Tarvittavat resurssit olivat saatavilla ja työllä oli julkisen hallinnon tuki. Yhdyskunta- ja kaupunkisuunnittelun yksikön kehittämä ja ylläpitämä Espoon Keskusten kehittämisfoorumi oli erinomainen alusta osallistavan palvelun kehitykseen ja paikallistietämyksen analysointi- ja visualisointimenetelmien kokeilemiseen.

Palvelun todettiin täyttävän vaatimukset kaupunkisuunnittelun ja osallistavuuden edistämiseksi. Sen hyödyllisyyttä päätöksenteon tukena ei voitu kuitenkaan osoittaa ennen kuin sen varsinaisesta käytöstä on kerätty kokemuksia. Hyödyllisyyden osoittaminen edellyttää työtä oikeanlaisten mittareiden löytämiseksi, koska yksittäisen muutoksen vaikutuksia osallistavuuteen ja päätöksentekoon on vaikea mitata.

Palvelu on yhteensopiva Kehittämisfoorumin kanssa ja noudattaa järjestelmän dokumentaatiossa annettuja linjauksia sivuston luonteesta. Palvelun tekninen toteutus

sujui mallikkaasti, tosin siihen kului hieman arvioitua enemmän aikaa. Palvelu voitiin toteuttaa siten, että se voidaan ottaa käyttöön myös tulevilla Kehitysympäristöissä tai muissa vastaavissa järjestelmissä.

Tekstianalyysin toteutuksessa käytetty latentti semanttinen indeksointi osoittautui hyväksi menetelmäksi vapaamuotoisen kommenttiaineiston semantiikan tulkitsemiseen. Kommenttien puheenaiheet voitiin esittää kartalla mielekkäällä tavalla Voronoi-diagrammia hyödyntäen.

Suunnitellun palvelukokonaisuuden muut palvelut voidaan kehittää hyödyntäen toteutetun palvelun tekniikkaa ja tuloksia. Ennen kaikkea työn tulokset osoittavat, että kaupunkisuunnittelua ja osallistavuutta voidaan kehittää tilastollisen tekstianalyysin ja visualisoinnin menetelmillä. Työssä käytiin läpi erilaisia vaihtoehtoja puheenaiheiden analysoinnin toteuttamiseen paikkatietojärjestelmässä, eikä ole syytä olettaa, että valittu menetelmä olisi ainoa tapa tämän tyyppisen palvelun toteuttamiseksi.

Menetelmä on ohjaamaton ja käytettävän aineiston ja kontekstin suhteen itsenäinen. Siinä ei juurikaan hyödynnetä kaupunkisuunnittelun erityispiirteitä. Toimintaan ei esimerkiksi ohjelmoitu erityiskäsittelyä kaupunkisuunnitteluun liittyvän termistön suhteen, eivätkä paikkatietojärjestelmässä olevat kaavahankkeet vaikuta analyysin tulokseen millään tavalla. Sen vuoksi sen toiminta ei rajoitu vain kaupunkisuunnittelun tapaukseen, vaan sitä on syytä kokeilla yhtä lailla myös muilla aloilla kuten ympäristönsuojelussa, luonnonvarojen hallinnassa tai liikennetutkimuksessa toimivien osallistavien verkkopalveluiden kehittämiseen.

## 8 Yhteenveto

Paikallistietämystä voitaisiin hyödyntää nykyistä enemmän kaupunkisuunnitteluun liittyvän päätöksenteon tietolähteenä. Paikallistietämystä voidaan kerätä osallistamalla paikalliset toimijat mukaan suunnitteluprosessiin esimerkiksi osallistavan GIS:n avulla. Osallistamiskäytännön kehittäminen vaatii kuitenkin perehtymistä sekä teknisiin että sosiaalisiin seikkoihin.

Keskeistä on tapa, jolla tietoa käsitellään paikkatietojärjestelmässä ja jolla se välitetään järjestelmän käyttäjille. Paikallistietämyksen visualisointi paikkatietojärjestelmän kartalla on tällä hetkellä melko tutkimaton aihepiiri, vaikka tarve sille on suuri. Visualisoimalla osallisten antamien kommenttien aihepiirejä paikkatietojärjestelmän kartalla saadaan suuri määrä tietoa esitettyä tehokkaasti sekä osallisten omaan että asiantuntijoiden ja päättäjien käyttöön.

Työn tavoitteena oli tutkia, voidaanko kansalaisten osallistamista ja kaupunkisuunnittelun prosessia parantaa paikkatietojärjestelmässä toteutetun tilastollisen tekstianalyysipalvelun avulla. Tutkimus tehtiin suunnittelemalla ja toteuttamalla verkkopalvelu, jonka avulla voidaan analysoida paikallisten toimijoiden antamien kommenttien puheenaiheita ja visualisoida analyysin tulos paikkatietojärjestelmän kartalla. Tekstianalyysi toteutettiin latentilla semanttisella indeksoinnilla ja tuloksen visualisoinnissa hyödynnettiin Voronoi-diagrammia. Tulokset olivat pääosin positiivisia ja asetetut tavoitteet voitiin saavuttaa. Palvelun käytöstä saatavan kokemuksen myötä voidaan jatkossa analysoida sen vaikutuksia kaupunkisuunnittelun päätöksentekoprosessiin sekä kansalaisten osallistamiseen.

Vaikka kehityksen alustana toimiva Espoon Keskuksen Kehittämisfoorumi ei uuden analyysipalvelun myötä astuisi osallistavuuden portaikossa seuraavalle askelmalle, se ainakin vahvistaa asemaansa tehokkaana osallisten kuulemisen foorumina. Kehitetty analyysimenetelmä ei ole vain kaupunkisuunnittelun apuväline, vaan ennen kaikkea toimiva työkalu laadullisen paikkatiedon analysointiin.



## Lähdeluettelo

Abbot, J. & Chambers, R. & Dunn, C. & Harris, T. & de Merode, Emmanuel & Porter, G. & Townsend, J. & Weiner, D. 1998. *Participatory GIS: opportunity or oxymoron?* Participatory learning & action PLA notes. Vol. 33. S. 27-34.

Ahas, R. & Mark, Ü. 2005. *Location based services – new challenges for planning and public administration?* Futures. Vol. 37. S. 547-561. ISSN 0016-3287 DOI 10.1016/j.futures.2004.10.012.

Al-Kodmany, K. 2002. *Visualization tools and methods in community planning: from freehand sketches to virtual reality.* Journal of planning literature. Vol. 17:2. S. 189-211. DOI 10.1177/088541202762475946.

Ambler, S. W. 2004. *The object primer: agile model-driven development with UML 2.0.* Cambirdge University press. 572 s. ISBN 0-521-54018-6.

Ambler, S. W. 2005. *The elements of UML 2.0.* Cambirdge University press. 200 s. ISBN 0-521-61678-6.

Anon. 2010. *Convex and concave polygons.* Wikipedia. [Vapaa sanakirja] [Viitattu: 28.1.2010]. Saatavissa: [http://en.wikipedia.org/wiki/Convex\\_polygon](http://en.wikipedia.org/wiki/Convex_polygon).

Appleton, K. & Lovett, A. 2005. *GIS-based visualization of development proposals: reactions from planning and related professionals.* Computers, environment and urban systems. Vol. 29. S. 321-339. ISSN 0198-9715.

Arnstein, S. R. 1969. *A ladder of citizen participation.* Journal of the American institute of planners. Vol. 35:4. S. 216-224.

Arterton, F. C. 1988. *Political participation and “teledemocracy”.* Political science and politics. Vol. 21 :3. S. 620-627.

Aurigi, A. 2000. *Digital city or urban simulator?* Digital cities. LNCS 1765. S. 33-44.

- Bertin, J. 1967 (1999). *Sémiologie Graphique*. Uusintapainos. Les Re-impressions des Editions de l'Ecole des Hautes Etudes En Sciences Sociales. 442 s. ISBN 2-7132-1277-4.
- Bingham, E. 2003. *Advances in independent component analysis with applications to data mining*. Väitöskirja. Teknillinen korkeakoulu, Tietotekniikan laitos, Informaatiotekniikan laboratorio. Espoo. ISBN 951-22-6819-1 (painettu) ISBN 951-22-6820-5 (sähköinen) ISSN 1459-7020.
- Bojórquez-Tapia, L. A. & Díaz-Mondragón, S. & Ezcurra, E. 2001. *GIS-based approach for participatory decision making and land suitability assessment*. International journal of geographical information science. Vol. 15:2. S. 129-151. ISSN: 1365-8816 (painettu) ISSN 1362-3087 (sähköinen) DOI 10.1080/13658810010005534.
- Booker, A. & Condliff, M. & Greaves, M. & Holt, F. B. & Kao, A. & Pierce, D. J. & Poteet, S. & Wu, Y-J. J. 1999. *Visualizing text data sets*. Computing in science & engineering. Vol. 1:4. S. 26-35. ISSN 1421-9615.
- Burke, J. & Estrin, D. & Hansen, M. & Parker, A. & Ramanathan, S. & Reddy, S. & Srivastava, M. B. 2006. *Participatory sensing*. WSW'06 at SenSys '06.
- Carpini, M. X. D. & Cook, F. L. & Jacobs, L. R. 2004. *Public deliberations, discursive participation and citizen engagement: a review of the empirical literature*. Annual review of political science. Vol. 7:1. S. 315-344. ISSN 1094-2939.
- Carver, S. 2003. *The future of participatory approaches using geographic information: developing a research agenda for the 21st century*. URISA Journal. Vol. 15. S. 61-71.
- Chew, L. P. 1990. *Building Voronoi diagrams for convex polygons in linear expected time*. Technical Report: PCS-TR90-147.
- Comon, P. 1992. *Independent component analysis*. Higher order statistics.
- Crampton, J. W. 2002. *Interactivity types in geographic visualization*. Cartography and geographic information science. ISSN 1523-0406 (painettu) ISSN 1545-0465

(sähköinen)

Danahy, J. 1999. *Visualization data needs in urban environmental planning and design*. Photogrammetric week „99“.

Delaney, B. 2000. *Visualization in urban planning: they didn't build LA in a day*. IEEE Computer graphics and applications. Vol. 20:3. S. 10-16. Doi 10.1109/38.844365.

Dragičević, S. & Balram, S. 2004. *A web GIS collaborative framework to structure and manage distributed planning processes*. Journal of geographical systems. Vol. 6. S. 133-153. DOI 10.1007/s10109-004-0130-7

Dunn, C. E. 2007. *Participatory GIS – a people's GIS?* Progress in human geography. Vol. 31:5. S. 616-637. DOI 10.1177/0309132507081493

Elwood, S. 2006. *Critical issues in participatory GIS: deconstructions, reconstructions, and new research directions*. Transactions in GIS. Vol. 10:5. S. 693-708.

Espoon kaupunki, palvelukeskus, Tieto- ja tutkimuspalvelut. 2001. *Espoon erilaiset alueet. Taustaselvitys Espoon alueellisesta segregoitumisesta*. Saatavissa: <http://www.espoo.fi/binary.asp?path=1;28;11894;37617;73021;73526;73528;73540&fileId=FileAttachment>.

Espoon Keskusten kehittämisfoorumin ylläpito. 2010. *Sivuston esittely*. [WWW-sivu]. [Viitattu 28.1.2010]. Saatavissa: <http://test.espoonkeskus.fi/node/433>.

Feng, L. & Hansen, L. K. 2006. *Phonemes as short time cognitive components*. IEEE international conference on acoustics, speech and signal processing. ICASSP 2006 Proceedings. Vol. 5. S. 14-19. ISBN 1-4244-0469-X

Fortune, S. 1992. *Voronoi diagrams and Delaunay triangulations*. Teoksessa: Du, Ding-Zu (toim.) & Hwang, Frank (toim.). Computing in Euclidean geometry. Singapore: World publications printers Pte. Ltd. S. 193-234. ISBN 981-02-0966-5.

Greene, D. & Cunningham, P. & Mayer, R. 2008. *Unsupervised learning and clustering*.

- Teoksessa: Cunningham, Pádraig (toim.). Machine learning techniques for multimedia. Case studies on organization and retrieval. Berlin, Germany: Springer-Verlag Berlin Heidelberg. S. 51-58. ISBN 978-3-540-75170-0.
- Hansen, L. K. & Ahrendt, P. & Larsen J. 2005. *Towards cognitive component analysis*. AKRR'05 – International and interdisciplinary conference on adaptive knowledge representation and reasoning. Pattern recognition society of Finland.
- Henrion, M. & Breese, J. S. & Horvitz, E. J. 1991. *Decision analysis and expert systems*. AI magazine. Vol. 12:4. S. 64-91. ISSN 0738-4602.
- Higgs, G. 2006. *Integrating multi-criteria techniques with geographical information systems in waste facility location to enhance public participation*. Waste manage res. Vol. 24. S. 105-117. DOI 10.1177/0734242X06063817.
- Inaba, K. & Katoh, N. & Imai, H. 1994. *Applications of weighted voronoi diagrams and randomization to variance-based k-clustering*. 10th computational geometry. Vol. 6. S. 332-339. ISBN 0-89791-648-4.
- Ingram, R. Benford, S. 1995. *Legibility enhancement for information visualisation*. Proceedings of the 6th IEEE visualization conference. S. 209-216. ISSN 1070-2385.
- Jolliffe, I. T. 2002. *Principal component analysis*. 2. painos. New York, USA: Springer-Verlag New York, Inc. 457 s. ISBN 0-387-95442-2.
- Jordan, G. 1999. *Public participation and GIS: report back*. PLA notes. Vol. 34. S. 16-17
- Kangas, J. & Store, R. 2003. *Internet and teledemocracy in participatory planning of natural resources management*. Landscape and urban planning. Vol. 62. S. 89-101. ISSN 0169-2046
- Kohonen, T. 1982. *The self-organizing map*. Neurocomputing. Vol. 21. S. 1-6. DOI 10.1016/S0925-2312(98)00030-7

Kyttä, M. & Kahila, M. 2007. *PehmoGIS: sähköistä viestintää asukkailta yhdyskuntasuunnittelijoille*. Teknillinen Korkeakoulu, Espoo. 22-23.11.2007. Maanmittaustieteiden päivät 2007 – Maanmittaus verkossa. Helsinki: Maanmittaustieteiden seura ry.

Lagus, H. & Honkela, T. & Kaski, S. & Kohonen, T. 1996. *Self-organizing maps of document collections: a new approach to interactive exploration*. KDD-96 Proceedings. ISBN 978-1-57735-418-5.

Lehtovirta, O. & Ahlqvist, T. & Loikkanen, T. & Eivola, A. 2010 *TEKBARO2010 Teknologiabarometri kansalaisten asenteista ja kansakunnan suuntautumisesta tietoon perustuaan yhteiskuntaan*. Helsinki: Forssan Kirjapaino Oy. 99 s.

Liu, B. & Hu, M. & Cheng, J. 2005. *Opinion observer: analyzing and comparing opinions on the web*. Proceedings of the 14th international conference on World Wide Web. S. 342-351. ISBN 1-59593-046-9.

*Maankäyttö- ja rakennuslaki 132/1999, 62§.*

*Maankäyttö- ja rakennuslaki 132/1999, 65§.*

MacEachren, A. M. 1992. *Visualizing uncertain information*. Cartographic perspective. Vol. 13. S. 10-19. ISSN 1048-9085.

MacEachren, A. M. 2001. *An evolving cognitive-semiotic approach to geographic visualization and knowledge construction*. Cartography and geographic information systems. Vol. 19. S. 197-200.

McCall, M. K. & Minang, P. A. 2005. *Assessing participatory GIS for community-based natural resources management: claiming community forests in Cameroon*. The geographical journal. Vol 171:4. S. 340-356. ISSN 0016-7398.

McCall, M. K. 2002. *Seeking good governance in participatory-GIS: a review of processes and governance dimensions in applying GIS to participatory spatial planning*. Habitat International. Vol 27. S. 549-573. DOI 10.1016/S0197-3975(03)00005-5

- McCall, M. K. 2004. *Can participatory-GIS strengthen local-level spatial planning? Suggestions for better practice*. Teoksessa: McCall, M. K. GISDECO 2004. Skudai, Johor, Malaysia. 10-12.5.2004. GISDECO.
- Mino, E. 2000. *Experiences of European digital cities*. Digital cities. S. 58-72.
- Miranda, A. A. & Le Borgne, Y-A. & Bontempi, G. 2008. *New routes from minimal approximation error to principal components*. Neural process lett. Vol. 26. S. 197-207. DOI 10.1007/s11063-007-9069-2.
- Murdock, B. B., jr. 1962. *The serial position effect in free recall*. Journal of experimental psychology. Vol. 64:5. S. 482-488.
- Nummi, P. 2007. *Verkko-osallistuminen kaavoituksen tukena*. Diplomityö. Teknillinen korkeakoulu, Arkkitehtuurin laitos. 127 s.
- OPUS Oppiva kaupunkisuunnittelu ja asumisen arki*. 2010. [WWW-sivusto]. [Viitattu 28.1.2010]. Saatavissa: <http://opus.tkk.fi/>
- Pain, R. 2004. *Social geography: participatory research*. Progress in human geography. Vol. 28:5. S. 652-663. DOI 10.1191/030913254ph511pr.
- Popping, R. 2000. *Computer-assisted text analysis*. London, UK: SAGE Publications Ltd. 206 s. ISBN 0-7619-5379-5.
- Rantanen, H. 2008. *Espoonatori? Kehittämisfoorumin keskustelualue*. [Keskustelufoorumi] Espoon Keskusten Kehittämisfoorumin keskustelualue. 11.9.2008. [Viitattu 28.1.2010]. Saatavissa: <http://www.espoonkeskus.fi/keskustelu/index.php?action=printpage;topic=133.0>.
- Rhyne, T. M. 1997. *Going virtual with geographic information and scientific visualization*. Computers & Geosciences. Vol. 23:4. S. 489-491. ISSN 0098-3004.
- Ruusula, M. 2001. *Palveleeko verkko kuntalaisia*. Helsinki: Julkisen verkkoasioinnin kehittämishanke (JUNA). 46 s. ISBN 951-734-467-8.

Saward, M. 1994. *Democratic theory and indices of democratization*. Teoksessa: Beetham, David (toim.) *Defining and measuring democracy*. London, UK: SAGE publications Ltd. S. 6-23. ISBN 0-8039-7789-1.

Scrutiny of acts and regulations committee (Parliament of Victoria, Australia). 2003. *Inquiry into electronic democracy*. Government printer for the state of Victoria. Vol. 126. ISBN 978-0-7311-3038-X

Staffans, A. 2004. *Vaikuttavat asukkaat. Vuorovaikutus ja paikallinen tieto kaupunkisuunnittelun haasteina*. Väitöskirja. Teknillinen korkeakoulu, Yhdyskuntasuunnittelun tutkimus- ja koulutuskeskus. Helsinki: Yliopistopaino Oy. 312 s. ISBN 951-22-6962-7 ISSN 1455-7789.

STT-IS. 2007. *Naapuri kuuli itkua ja huutoa aamuyöllä*. Ilta-Sanomat. [Verkkolehti]. 7.8.2007. [Viitattu 28.1.2010]. Saatavissa: <http://www.iltasanomat.fi/uutiset/kotimaa/uutinen.asp?id=1417271>.

Sui, D. Z. 1997. *Reconstructing urban reality: from GIS to electropolis*. Urban geography. Vol. 18:1. ISSN 0272-3638.

Suominen, O. 2008. *Käyttäjäkeskeinen moninäkömähaku semanttisessa portaalissa*. Pro gradu-tutkielma. Helsingin yliopisto, Tietojenkäsittelytieteen laitos. 66 s. Saatavissa: <http://www.seco.tkk.fi/publications/2008/suominen-gradu-2008.pdf>.

Tang, K. X. & Waters, N. M. 2005. *The internet, GIS and public participation in transportation planning*. Progress in planning. Vol. 64. S. 7-62. DOI 10.1016/j.progress.2005.03.004.

Tress, B. & Tress, G. 2003. *Scenario visualization for participatory landscape planning – a study from Denmark*. Landscape and urban planning. Vol. 64. S. 161-178. ISSN 0169-2046.

Tripathi, N. & Bhattarya, S. 2004. *Integrating indigeous knowledge and GIS for participatory natural resource management: state-of-the-practice*. The electronic journal on information systems in developing countries. Vol. 7:3. S. 1-13. ISSN 1681-

4835.

Wendy, L. & Sundheim, B. 1991. *A performance evaluation of text-analysis technologies*. AI magazine. Vol. 12:3. S. 81-94. ISSN 0738-4602.

Voss, A. & Denisovich, I. & Gatalsky, P. & Gavouchidis, K. & Klotz, A. & Roeder, S. & Voss, H. 2004. *Evolution of a participatory GIS*. Computers, environment and urban systems. Vol. 28. S. 635-651. ISSN 0198-9715 DOI 10.1016/j.compenvurbsys.2003.12.003.

Ylläpidon ohje kerätyn materiaalin analysoinnista. 2009.

*KerätynMateriaalinAnalysointi.doc*. [Tekstitiedosto]. Ei julkisesti saatavilla.

Zimmerman, M. A. & Rappaport, J. 1988. *Citizen participation, perceived control, and psychological empowerment*. American journal of community psychology. Vol. 16:5. S. 725-750. ISSN 0091-0562 (painettu) ISSN 1573-2770 (sähköinen) DOI 10-1007/BF00930023.



## Liite 1. Analyysipalvelun käyttötapaukset

*Taulukko 6, Käyttötapaus UC 1, Yksittäisen puheenaiheen sijainnin selvittäminen*

<b>Käyttötapaus</b>	UC 1
<b>Nimi</b>	Yksittäisen puheenaiheen sijainnin selvittäminen
<b>Toimija</b>	Kaupunkisuunnittelija tai kansalainen
<b>Esiehdot</b>	Käyttäjä on siirtynyt paikallistietokarttanäkymän analysoivälilehdelle. Käyttäjällä on tiedossa puheenaihe, jonka sijainnin haluaa selvittää.
<b>Kuvaus</b>	1. Käyttäjä kirjoittaa analyysipalvelun tekstikenttään puheenaiheen, painaa käynnistysnappia ja odottaa tulosta. 2. Käyttäjä katsoo kartalta, missä paikoissa puheenaihe esiintyy tarkastelemalla värin peittämää aluetta kartalla.
<b>Jälkiehdot</b>	Käyttäjällä on näkemys puheenaiheen maantieteellisestä esiintymisestä.
<b>Erityisvaatimukset</b>	Käyttäjä ymmärtää analyysipalvelun toiminnan idean ja osaa tulkita kartalla esitettyä abstraktia visuaalista informaatiota.
<b>Muuta huomiotavaa</b>	

*Taulukko 7, Käyttötapaus UC 2, Useamman puheenaiheen keskinäinen maantieteellinen vertailu*

<b>Käyttötapaus</b>	UC 2
<b>Nimi</b>	Useamman puheenaiheen keskinäinen maantieteellinen vertailu
<b>Toimija</b>	Kaupunkisuunnittelija
<b>Esiehdot</b>	Käyttäjä on siirtynyt paikallistietokarttanäkymän analysoivälilehdelle. Käyttäjällä on tiedossa puheenaiheet, joita haluaa vertailla.
<b>Kuvaus</b>	1. Käyttäjä kirjoittaa analyysipalvelun tekstikenttään yhden puheenaiheen, painaa käynnistysnappia ja odottaa tulosta. 2. Käyttäjä kirjoittaa samaan tekstikenttään toisen puheenaiheen ja painaa käynnistysnappia ja odottaa tulosta. 3. Käyttäjä katsoo kartalta, missä paikoissa puheenaiheet

	esiintyvät yhdessä ja missä erikseen tarkastelemalla eri värien peittämiä alueita kartalla.
<b>Jälkiehdot</b>	Käyttäjällä on näkemys puheenaiheiden maantieteellisestä yhteisesiintymisestä.
<b>Erityisvaatimukset</b>	Käyttäjä ymmärtää analyysipalvelun toiminnan idean ja osaa tulkita kartalla esitettyä abstraktia visuaalista informaatiota.
<b>Muuta huomioitavaa</b>	

*Taulukko 8, Käyttötapa UC 3, Omaa kommenttia vastaavan puheenaiheen sijainnin selvittäminen*

<b>Käyttötapa</b>	UC 3
<b>Nimi</b>	Omaa kommenttia vastaavan puheenaiheen sijainnin selvittäminen
<b>Toimija</b>	Kansalainen
<b>Esiehdot</b>	Käyttäjä on siirtynyt paikallistietokarttanäkymään. Käyttäjällä on jättänyt kommentin paikallistietokarttaan.
<b>Kuvaus</b>	1. Käyttäjä etsii kartalta jättämänsä kommentin ja klikkaa sitä hiirellä. 2. Käyttäjä kopioi kommenttinsa tekstisisällön leikepöydälle. 3. Käyttäjä liittää tekstisisällön analyysipalvelun hakukenttään. ja painaa käynnistysnappia. 4. Käyttäjä katsoo kartalta, missä paikoissa kommenttia vastaava puheenaihe esiintyy tarkastelemalla värin peittämää aluetta kartalla.
<b>Jälkiehdot</b>	Käyttäjä tietää, miten hänen oma kommenttinsa liittyy käytyyn keskusteluun, ja missä paikoissa sama asia on puhuttanut muita käyttäjiä.
<b>Erityisvaatimukset</b>	Käyttäjä ymmärtää analyysipalvelun toiminnan idean ja osaa tulkita kartalla esitettyä abstraktia visuaalista informaatiota.
<b>Muuta huomioitavaa</b>	

*Taulukko 9, Käyttötapa UC 4, Väliaikaistiedostojen poistaminen*

<b>Käyttötapa</b>	UC 4
<b>Nimi</b>	Väliaikaistiedostojen poistaminen
<b>Toimija</b>	Ylläpitäjä
<b>Esiehdot</b>	Käyttäjällä on käyttöoikeus palvelimen tiedostoihin.  Käyttäjä on kirjautunut sisään palvelimen hakemistojärjestelmään esimerkiksi etäpääteohjelmalla ja siirtynyt hakemistoon, jossa on analyysipalvelun moduliin liittyvät tiedostot.
<b>Kuvaus</b>	1. Käyttäjä siirtyy alihakemistoon temp.  2. Käyttäjä tyhjentää hakemiston poistamalla sieltä kaikki tiedostot.
<b>Jälkiehdot</b>	Palvelimelta on siivottu turhat väliaikaistiedostot pois.
<b>Erityisvaatimukset</b>	
<b>Muuta huomioitavaa</b>	Palvelu luo temp-hakemistoon tekstimuotoisia väliaikaistiedostoja joka kerta kun palvelua käytetään. Tiedostot eivät ole kovinkaan suuria, mutta ajan myötä niitä voi kertyä paljon. Tiedostoja ei kannata poistaa turhaan, koska niistä voi olla apua ongelmatilanteiden selvittelyssä.

*Taulukko 10, Käyttötapa UC 5, Palvelun käyttöönotto uudessa ympäristössä*

<b>Käyttötapa</b>	UC 5
<b>Nimi</b>	<i>Palvelun käyttöönotto uudessa ympäristössä</i>
<b>Toimija</b>	Ylläpitäjä
<b>Esiehdot</b>	Käyttäjällä on käyttöoikeus palvelimen tiedostoihin.  Käyttäjällä on sivuston täysi ylläpito-oikeus.
<b>Kuvaus</b>	Palvelun käyttöönotosta uudessa ympäristössä on olemassa erillinen ohje, jota seuraamalla se tulee tehdä.
<b>Jälkiehdot</b>	Palvelu on käytössä uudessa ympäristössä.
<b>Erityisvaatimukset</b>	
<b>Muuta huomioitavaa</b>	